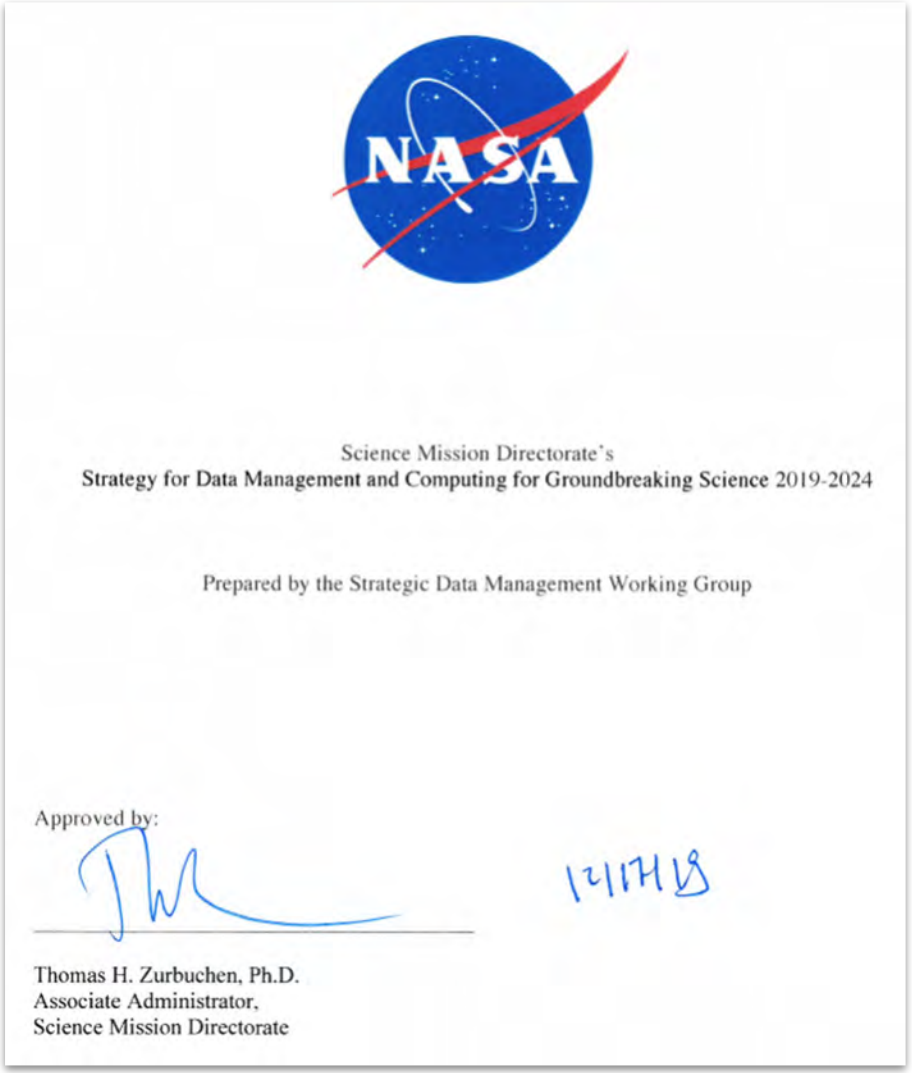
ESO Processing Workshop #1

Expanding participation, improving reproducibility, and accelerating scientific discovery for societal benefit.

October 14, 2021

Kevin Murphy

SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024



Science Mission Directorate's
Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

Prepared by the Strategic Data Management Working Group

Approved by:

12/17/19

Thomas H. Zurbuchen, Ph.D.
Associate Administrator,
Science Mission Directorate

Vision: To enable **transformational open science** through continuous evolution of science data and computing systems for NASA's Science Mission Directorate.

Mission:

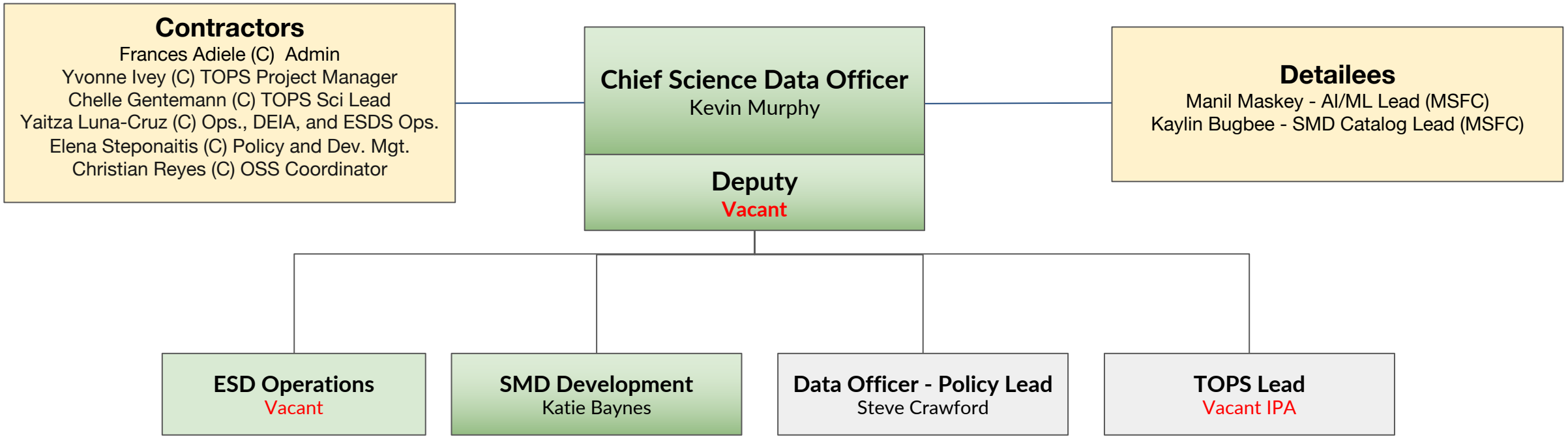
- Lead an **innovative and sustainable program** supporting NASA's unique science missions with academic, international and commercial partners to **enable groundbreaking discoveries with open science**.
- **Continually evolve systems** to ensure they are usable and support the latest analysis techniques while protecting scientific integrity.

Goal 1: Develop and Implement Capabilities to Enable Open Science

Goal 2: Continuous Evolution of Data and Computing Systems

Goal 3: Harness the Community and Strategic Partnerships for Innovation

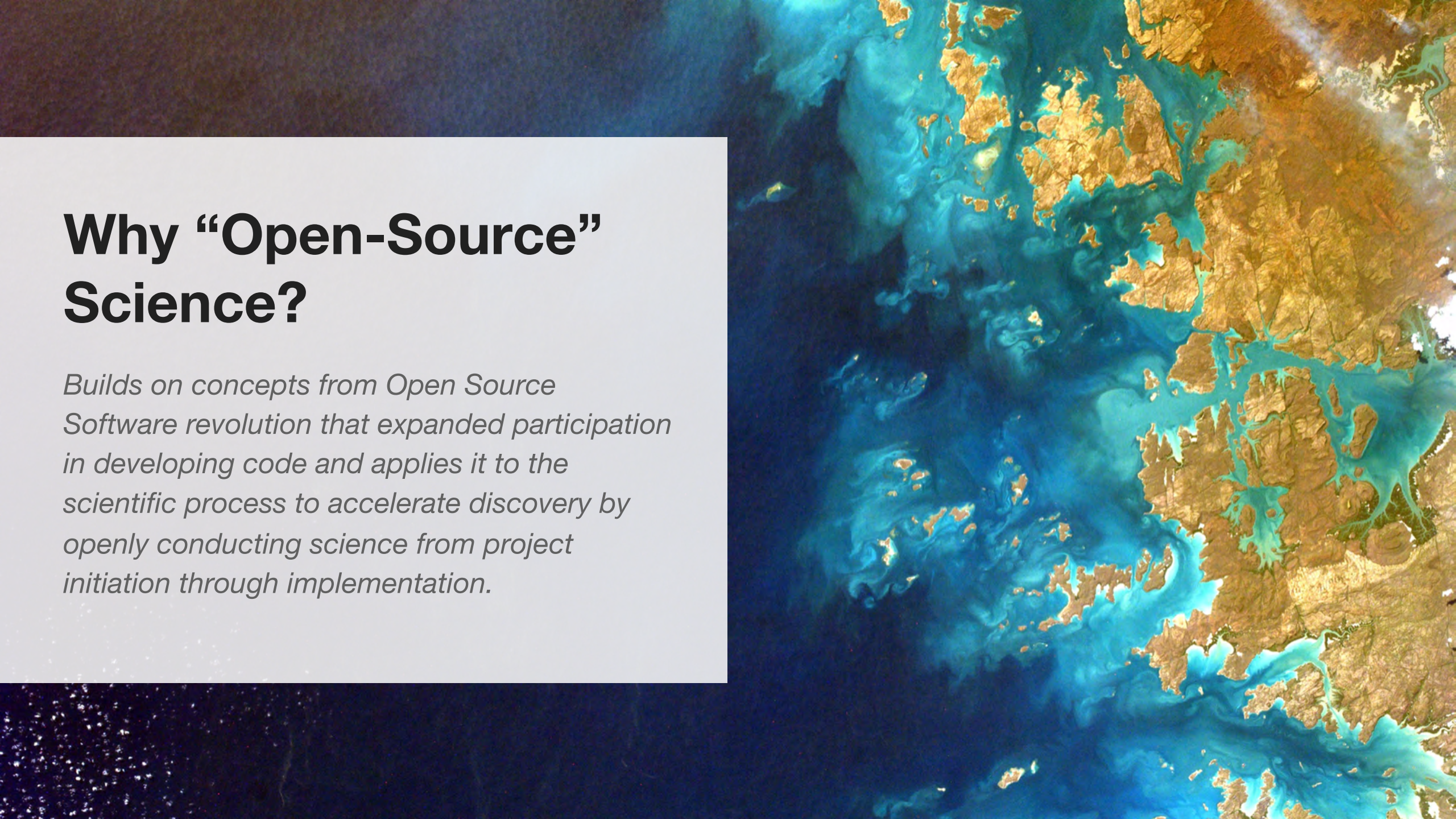
Chief Science Data Office



Open Science

"We define open science as a collaborative culture enabled by technology that empowers the **open sharing of data, information, and knowledge** within the scientific community and the wider public to accelerate scientific research and understanding."

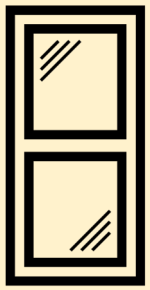
Ramachandran, R., Bugbee, K., & Murphy, K. J. Moving from Open Data to Open Science. Earth and Space Science, Wiley Publication https://doi.org/10.1029/2020EA001562



Why “Open-Source” Science?

Builds on concepts from Open Source Software revolution that expanded participation in developing code and applies it to the scientific process to accelerate discovery by openly conducting science from project initiation through implementation.

Building **trust in the scientific process** through **transparency, accessibility, inclusivity, and reproducibility**

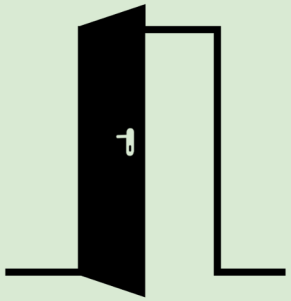


Open (**Transparent**) Science

Both the scientific process and results should be visible, accessible and understandable.

Open (**Accessible**) Science

Data, tools, software, documentation, publications should be accessible to all (FAIR).



Open (**Inclusive**) Science

The process and participants should welcome participation by and collaboration with diverse people and organizations.

Created by Gregor Cresnar from Noun Project

Open (**Reproducible**) Science

The scientific process and results should be open such that they are reproducible by members of the community.

Open-Source Science @ SMD

Initiates **Transform to OPen Science (TOPS),** a 5-year program to increase understanding and adoption of open science principles and techniques
Designates **2023 as Year of Open Science**

Continues **investments in open-source science digital infrastructure**, **cross-divisional AI capabilities and Digital Transformation activities**. (ROSES elements, data catalog, open journal database)

Prototype **common data catalog** by FY22Q4, ADS expansion

Initial investments in cross-division **open scientific cloud environments and data analysis platform prototypes**.

Divisional investments in Open-Source Science are **aligned** with this program.

Fiscal year	OSS Total ($M)
FY21	$8
FY22	$21
FY23	$20
FY24	$20
FY25	$20
FY26	$20
FY27	$20

Transform to OPen Science (TOPS) is a 5-year effort focused on capacity building, partner engagement, and incentives to help accelerate scientific discovery through open science.

TOPS Focus Areas

Public Engagement	• **Designate 2023 as Year of Open Science** • Partnering with professional orgs., publishing TOPS articles in high-impact journals • Engage early with historically excluded communities • TOPS GitHub
Capacity Building	• Create FAIR - Analysis-Ready Cloud-Optimized (ARCO) data • Develop learning resources • TOPS JupyterHub • Host and sponsor events (summer schools, multi-day trainings, massive open online courses)
Incentives	• Develop NASA Open Source Science Awards program • Leverage prizes and challenges and cross-division science use cases • Increased citizen science activities

Open-Source Science Policy for Earth System Observatory

A. **All mission data, metadata, software, databases, publications, and documentation shall be available on a full, free, open, and unrestricted basis starting in Phase B** with no period of exclusive access.
B. **Science workshops and meetings** shall be **open** to broad participation and documented in public repositories.

1	**Software shall be developed openly** in a publicly accessible, version-controlled platform using a **permissive software license allowing for community use and contributions.**	4	Scientific data, metadata, software, publications and documentation **shall be archived and made available by NASA and/or [Partner] starting in Phase B.**
2	**Manuscripts shall be published with open access licenses**; versions of as-accepted manuscripts shall be made available as open preprints and deposited in a NASA or [Partner] **repository upon publication.**	5	**NASA and [Partner]** software, documentation and data **shall be properly marked, cited, and/or attributed**. Metrics to measure and acknowledge open-source science contributions will be developed.
3	All mission **data, calibration information, and simulated products supporting development and validation of algorithms shall be made available without any conditions to use.**	6	**NASA and [Partner] will mutually develop an Open-Source Science Plan** that specifies details of collaboration.

* Projects should release all information with open licenses unless exceptions are granted based on laws or regulations, including classified, ITAR, EAR and CUI restrictions. CSDO evaluates and approves or declines deviation requests by projects for NASA.

Meeting Objective

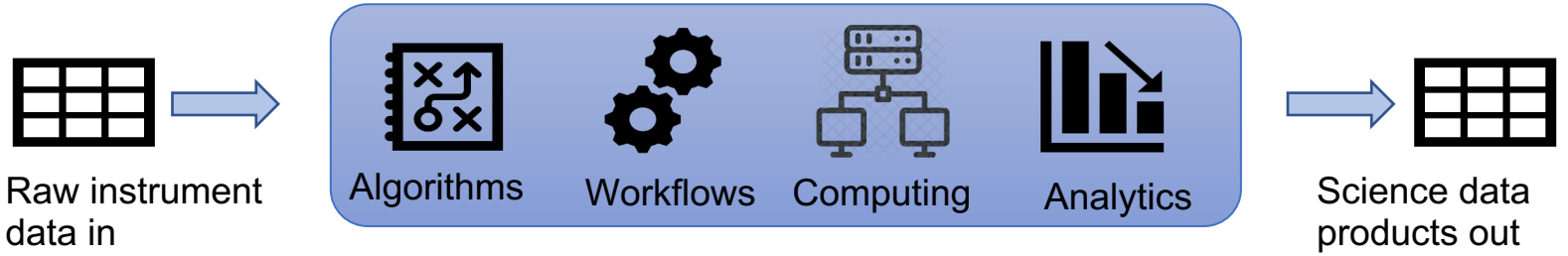
Open-Sourced Science Processing Study Goal

Why are we focuses on the data system?

Access to algorithms, workflows, computing, and analytics has been a major barrier to participating in NASA science. Opening it up provides greater opportunities for more people to participate in NASA science.

Identify and assess potential architectures that can meet the ESO mission science data processing objectives

- Promote open science principles,
- Enable data system efficiencies,
- Support Earth system science, and
- Expand participation in mission science beyond the funded science teams.



Study Principles

This is an open

EARTH SCIENCE
DATA SYSTEMS

NASA

Meeting goal

SPD-41: Science Information Policy

NASA SMD Scientific Information Policy

Steve Crawford
Science Data Officer

Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

Goal 1: Develop and Implement Capabilities to Enable Open Science		**Goal 2**: Continuous Evolution of Data and Computing Systems		**Goal 3:** Harness the Community and Strategic Partnerships for Innovation	
1.1	Develop and implement a **consistent open data and software policy** tailored for SMD	**2.1**	Establish **standardized approaches for all new missions** and sponsored research that encourage the adoption of advanced techniques	**3.1**	Develop **community of practice and standards group**
1.2	Upgrade capabilities at existing archives to **support machine readable data access using open formats and data services**	**2.2**	Integrate investment decisions in High-End Computing with the strategic needs of the research communities	**3.2**	Partner with a**cademic, commercial, governmental and international organizations**
1.3	Develop and implement a SMD data catalog to support discovery and access to complex scientific data across divisions	**2.3**	Invest in capabilities to use commercial cloud environments for open science	**3.3**	Promote opportunities for continuous learning as the field evolves through collaboration
1.4	Increase transparency into how science data are being used through a free and open unified journal server	**2.4**	Invest in the tools and training necessary to enable breakthrough science through application of AI/ML		

NASA Scientific Information

The information produced as part of NASA's scientific research activities represents a significant public investment. NASA holds this information as a public trust to increase knowledge and serve the public good.

Types of information:

- Publications: Scientific and technical documents released through print, electronic, or alternative media.
- Data: Scientific information that can be stored digitally and accessed electronically.
- Software: Computer programs in both source and object code that provide users some degree of scientific utility or produce a scientific result or service.

Why an Information Policy?

- Openly and widely sharing NASA research to maximize benefit and reach of the information
- The Policy consolidates a wide range of applicable laws, guidance, and policies
- Provides accessible and transparent policy for our program officers, scientists, and communities
- Helps support open science

Core Values in the Development of the policy

- Maximize the openness of useful scientific information produced by NASA SMD funding
- Preserve the scientific information produced by NASA SMD funding
- Support the community in accessing the information that they want
- Minimize the burden in complying with the policy

SPD-41: Scientific Information Policy

The science mission directorate has adopted SPD-41 that is a consolidation of existing policies applicable to SMD. These policies are based on our understanding of existing NASA and Federal guidance, and they are already part of solicitations for funding such as ROSES or SALMON Announcement of Opportunities. This applies to all SMD-funded activities related to producing scientific information, but the policy excludes restricted information.

- These policies are applicable to all current or future awards, contracts, or cooperative agreements for scientific activities.
- SPD-41: The Science Information Policy
- Science Information Policy Website

In addition, SMD will be releasing a Request for Information on proposed additions to the information policy based on new Federal guidance, NASA policy, National Academy studies, or community best practices.

Highlights from the General Policy

- All SMD-funded publications (publications funded by SMD or reporting on SMD-funded research) shall be made publicly accessible.
- SMD-funded data shall be made publicly available without fee or restriction of use.
- SMD-funded software should be released as open-source software.
- All SMD-funded activities shall have data management plans describing the management and release of data to facilitate the implementation of these information policies The DMP should include a description of the software to be used and how it will be managed.

Mission Specific Highlights to the Policy

- SMD shall commit to the full and open sharing of information produced by NASA SMD Missions. This includes observations, calibrations, coefficients, documentation, software, algorithms, technical reports, and any ancillary information or work product related to the Mission.
- There shall be no period of exclusive access to Mission data. A period after the data have been obtained may be allowed for activities such as calibration and validation of the data. This period shall be as short as practical and shall not exceed six months.

Research Specific Highlights to the Policy

- Research data shall become publicly available no later than the publication of the peer-reviewed article that describes it.
  - This includes data and software required to derive the findings communicated in figures, maps, and tables.
- In order to achieve reproducibility, research software developed using SMD funding and used in support of a scientific, peer-reviewed publication should be released as open source software no later than the publication date.

New Guidance and Additional Policies

Since 2015 there have been a range of new and proposed laws, recommendations, policies and Federal Guidance related to Open Science. We will be releasing an updated draft version of SPD-41 to include some of these changes.

Some divisions may enact these policies earlier, but these new policies will only be enacted once adopted and will fully apply only to new missions and investigations. Existing investigations should adopt the policy consistent with available resources.

TITLE II—OPEN GOVERNMENT DATA ACT

SEC. 201. SHORT TITLE.

This title may be cited as the "Open, Public, Electronic, and Necessary Government Data Act" or the "OPEN Government Data Act".

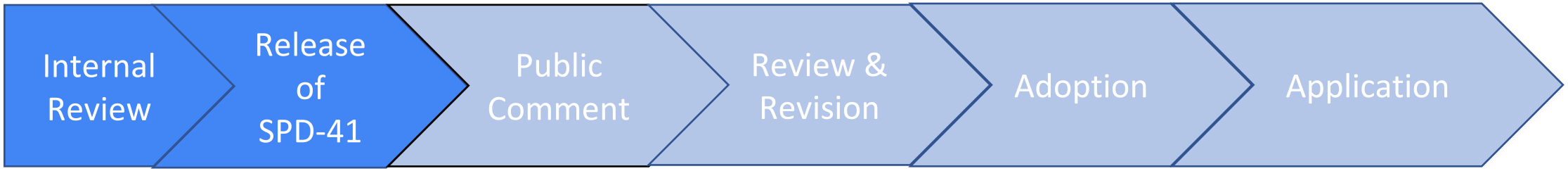
Highlights of Proposed Additions

- SMD-funded data should follow the FAIR Guiding Principles for scientific data management and stewardship. This means data should be findable, accessible, interoperable, and reusable (FAIR).
- Persistent identifier should be used for funding mechanisms and missions
- NASA data collections should have DOIs and meta-data available
- SMD-funded investigators should have a persistent identifier such as ORCID
- Mission software shall be developed openly allowing for community contributions

Highlights of Proposed Additions

- Scientifically useful data should be made publicly available at the end of the award
- SMD-funded software shall be released as open source software.
- Peer reviewed data and software shall be recognized as having the commensurate value as peer reviewed manuscripts.
- There will be further guidelines on compliance with the policy

Schedule for the policy development



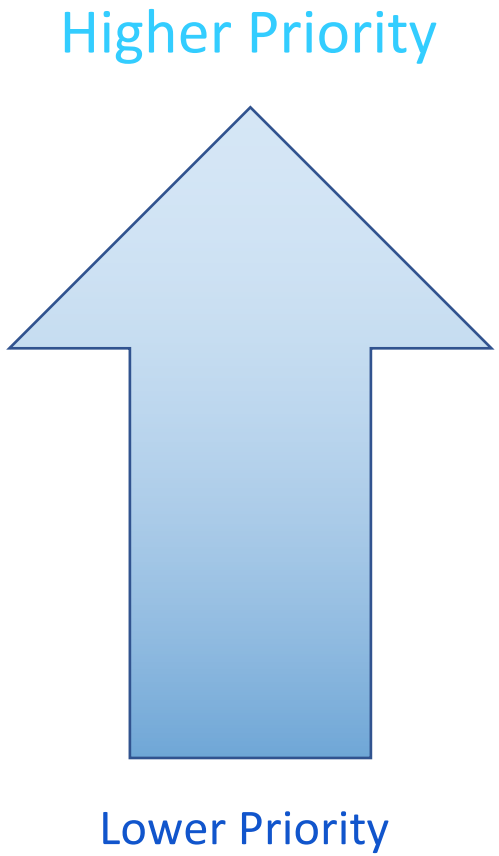
SPD-41 was released in September 2021. The schedule for the policy is:

- Approval of the proposed additions will occur no earlier than December 2022.
- If adopted, policy will be in place for ROSES and AO calls in 2023.
  - Some solicitations or divisions may adopt parts of the policy earlier.
  - Existing Missions and Researchers should adopt the policy as their resources allow.
- Except for variances, researchers funded from ROSES23 will need to be compliant.
  - Funded projects will typically start in 2024.
  - Funded publications will typically be produced starting in 2025

Next steps on the Information Policy

The development of the policy is only an early step in the overall process that will take place over the next 5 years.
Here are *some* of the next steps:

1. Identify ways to automate
2. Provide further guidance from the divisions
3. Provide training (TOPS)
4. Provide support for adopting open science
5. Support and/or develop technologies
6. Identify metrics for assessing compliance.



Open-Source Science Policy for Earth System Observatory

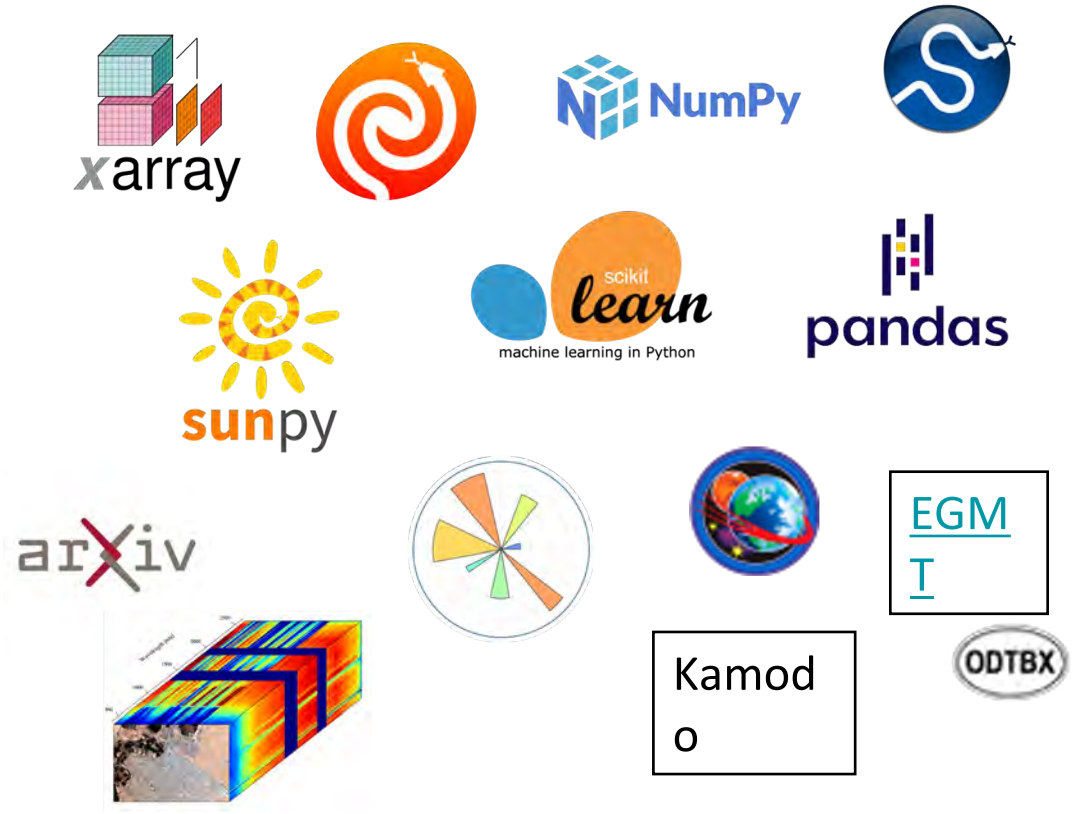
A. **All mission data, metadata, software, databases, publications, and documentation shall be available on a full, free, open, and unrestricted basis starting in Phase B** with no period of exclusive access.
B. **Science workshops and meetings** shall be **open** to broad participation and documented in public repositories.

1	**Software shall be developed openly** in a publicly accessible, version-controlled platform using a **permissive software license allowing for community use and contributions.**	4	Scientific data, metadata, software, publications and documentation **shall be archived and made available by NASA and/or [Partner] starting in Phase B.**
2	**Manuscripts shall be published with open access licenses**; versions of as-accepted manuscripts shall be made available as open preprints and deposited in a NASA or [Partner] **repository upon publication.**	5	**NASA and [Partner]** software, documentation and data **shall be properly marked, cited, and/or attributed**. Metrics to measure and acknowledge open-source science contributions will be developed.
3	All mission **data, calibration information, and simulated products supporting development and validation of algorithms shall be made available without any conditions to use.**	6	**NASA and [Partner] will mutually develop an Open-Source Science Plan** that specifies details of collaboration.

* Projects should release all information with open licenses unless exceptions are granted based on laws or regulations, including classified, ITAR, EAR and CUI restrictions. CSDO evaluates and approves or declines deviation requests by projects for NASA.

Supporting Open Source Software

- ROSES20 E.7 Open Source Tools, Frameworks, and Libraries selected 8 proposals supporting 14 different projects.
- ROSES21 F.8 Supplemental Open Source Software Awards - Support for existing grant holders to move to open source
- Improving the software release process for NASA civil servants and missions



Full description of supported projects is available on NSPIRES

- Working with STI/Software
- ADS
- STI DOIMS
- SMD Data catalog

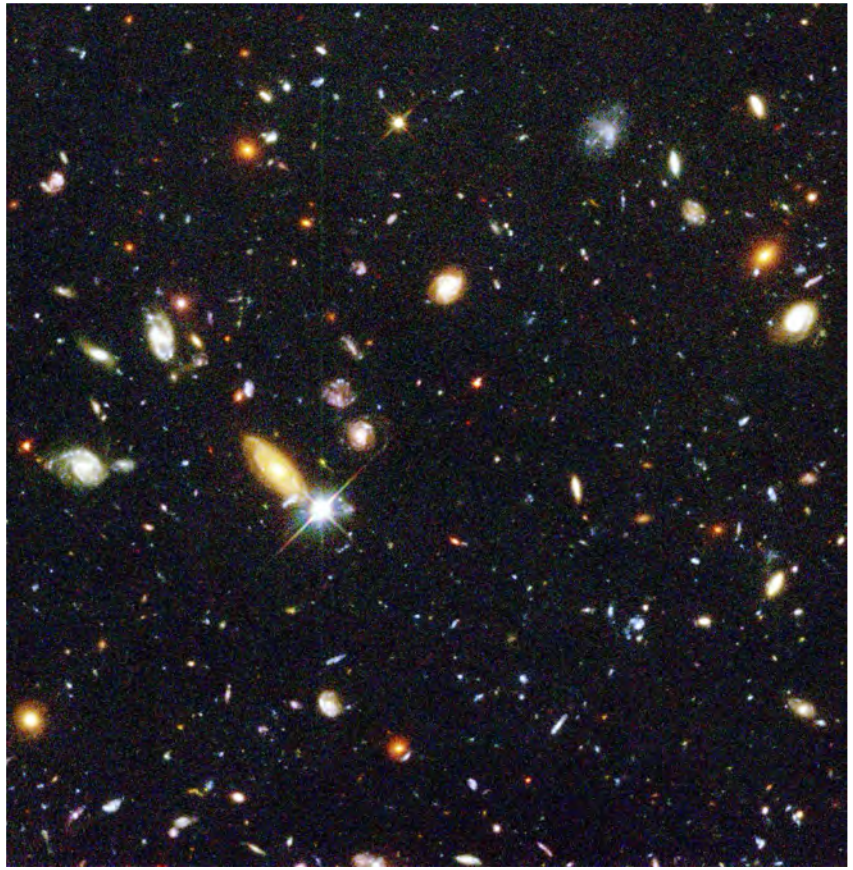
Increasing Access to SMD Publications

NASA has a mandate to ensure access to and reliable preservation of peer-reviewed publications that arise from NASA-funded research.

- NASA SMD supports publishing as Open Access and encourages making your publications available on preprint servers.
- SMD is funding ADS to expand its holdings in Heliophysics and Planetary Science.
- NASA STI is working on an agreement with CHORUS and improving the PubSpace interface.
- We will be working with our partners and publishers to further improve the process.
- We will be developing further guidance and services to make it automatic to preserve and make your publications accessible.

Making NASA's Data Accessible

NASA has a fantastic legacy in making our scientific data freely accessible and widely available. What are the next steps? How can we innovate?



Current OSSI Activities/Activities kicking off:

- SMD Data Catalog Search
- Standards for NASA Data
- AI Enable data
- Cloud ready data
- Openscapes

Some Open Questions

- What will we do to make research data accessible?
- How do we ensure that our data is accessible by all?
- How do we further empower citizen science?
- How do we make our data FAIR?
- How do we efficiently handle the large volumes of data and enable interoperability?
- How can we make use of the cloud and HEC?

Public Comment on the Policy

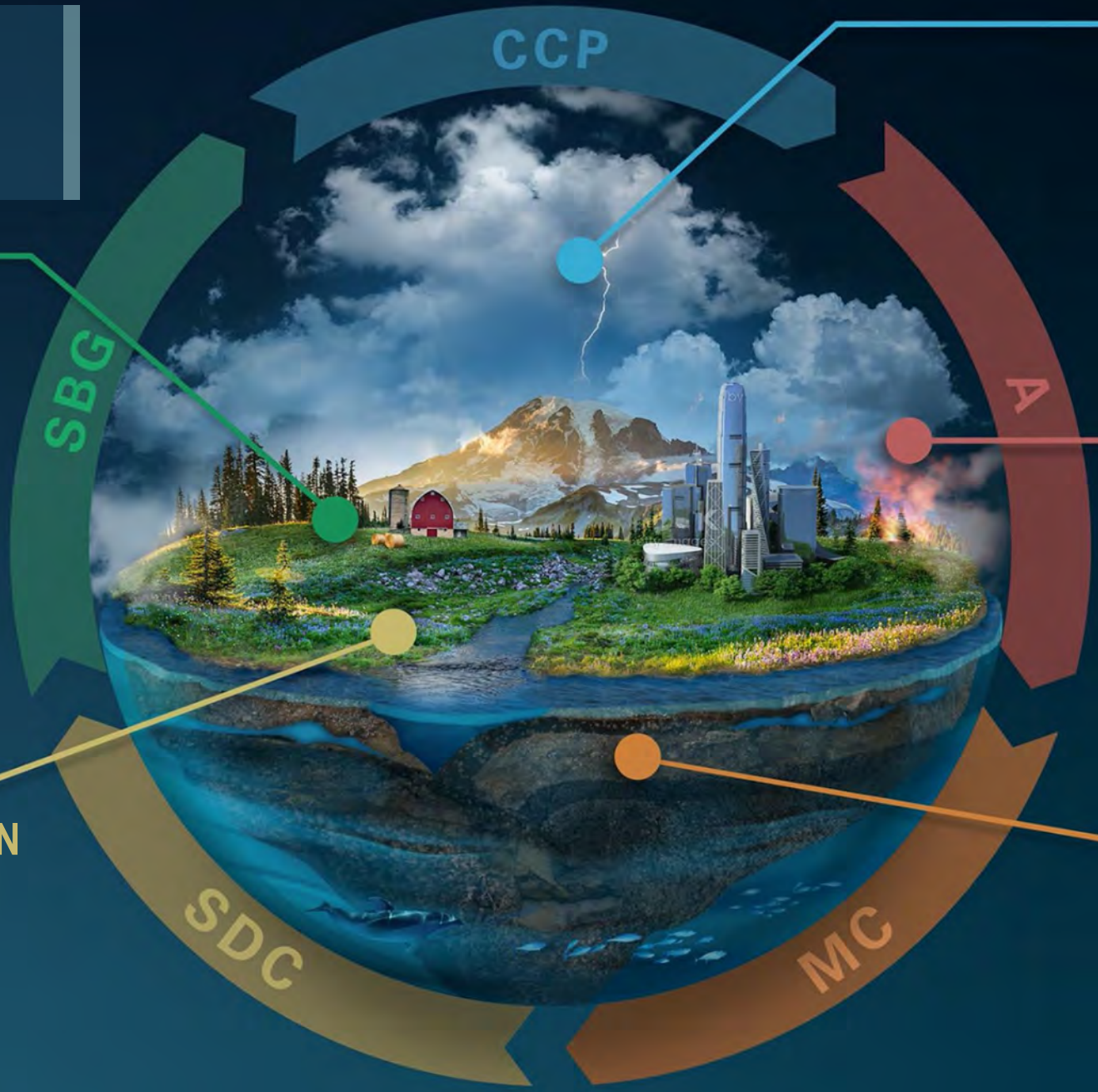
RFI to be released requesting information on:

- How will the proposed changes to the existing information policy impact the research activities of your communities?
- What support, services, training, funding, or further guidance is needed to support the successful implementation of the existing or proposed information policy?

Questions can also be sent to HQ-SMD-SPD41@mail.nasa.gov

Katie’s Talks

EARTH SYSTEM OBSERVATORY



CLOUDS, CONVECTION AND PRECIPITATION

Water and Energy in the Atmosphere

AEROSOLS

Particles in the Atmosphere

MASS CHANGE

Large-scale Mass Redistribution

SURFACE DEFORMATION AND CHANGE

Earth Surface Dynamics

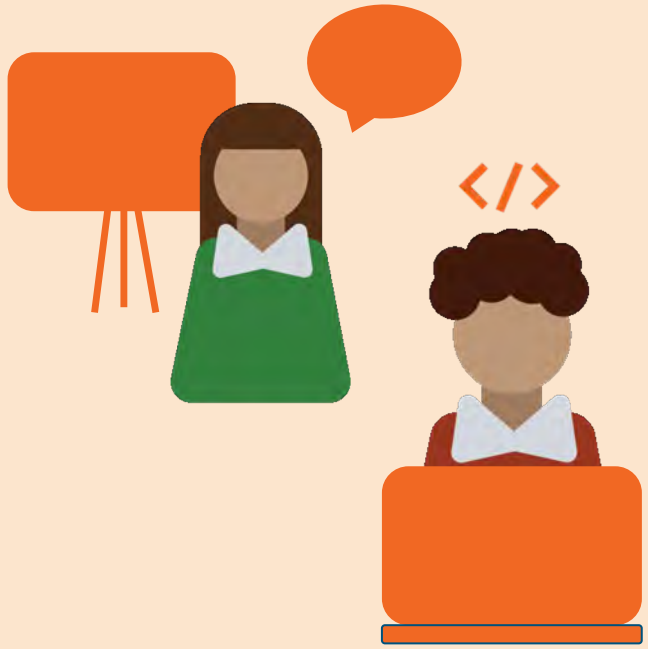
SURFACE BIOLOGY AND GEOLOGY

Earth Surface & Ecosystems

PROGRAM PRE-FORMULATION

PRE-PHASE A: Concept Studies

PROGRAM FORMULATION



PHASE A: Concept and Technology Development
PHASE B: Preliminary Design and Technology Completion

PROGRAM IMPLEMENTATION

PHASE C: Final Design and Fabrication
PHASE D: System Assembly, Integration and Test, Launch
PHASE E: Operations and Sustainment
PHASE F: Closeout

Open-Source Science Policy for Earth System Observatory

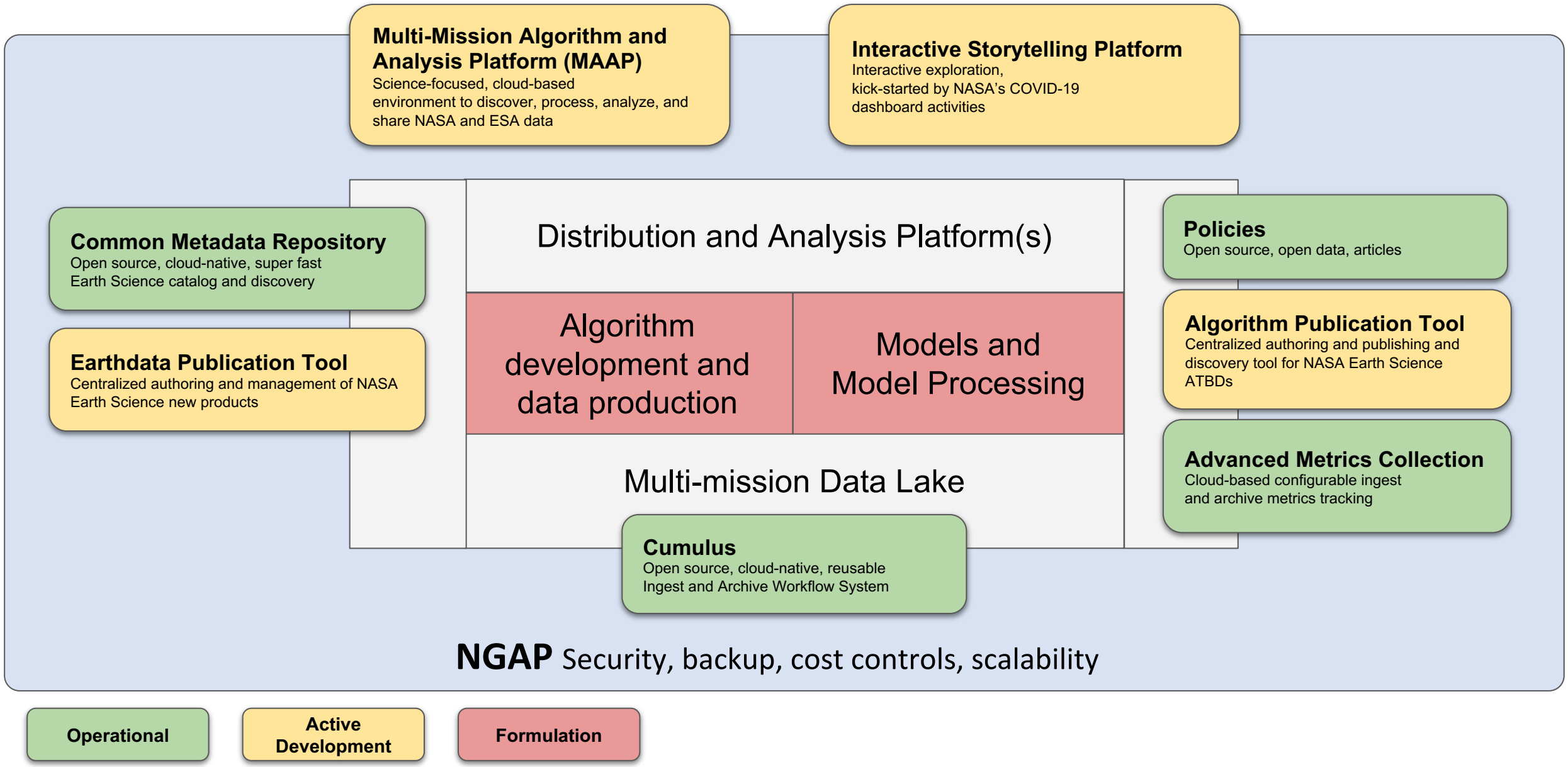
A. **All mission data, metadata, software, databases, publications, and documentation shall be available on a full, free, open, and unrestricted basis starting in Phase B** with no period of exclusive access.
B. **Science workshops and meetings** shall be **open** to broad participation and documented in public repositories.

1	**Software shall be developed openly** in a publicly accessible, version-controlled platform using a **permissive software license allowing for community use and contributions.**	4	Scientific data, metadata, software, publications and documentation **shall be archived and made available by NASA and/or [Partner] starting in Phase B.**
2	**Manuscripts shall be published with open access licenses**; versions of as-accepted manuscripts shall be made available as open preprints and deposited in a NASA or [Partner] **repository upon publication.**	5	**NASA and [Partner]** software, documentation and data **shall be properly marked, cited, and/or attributed**. Metrics to measure and acknowledge open-source science contributions will be developed.
3	All mission **data, calibration information, and simulated products supporting development and validation of algorithms shall be made available without any conditions to use.**	6	**NASA and [Partner] will mutually develop an Open-Source Science Plan** that specifies details of collaboration.

* Projects should release all information with open licenses unless exceptions are granted based on laws or regulations, including classified, ITAR, EAR and CUI restrictions. CSDO evaluates and approves or declines deviation requests by projects for NASA.

Envisioning DAACs as Science Enabling Centers

1	**Data and metadata stewardship** Ensuring **quality and fitness for purpose** of the organization's data and metadata assets
2	**Information management** Capturing and cataloging scientific information from publications into **searchable databases linked to data resources** Methods: semantic search, natural language processing, machine learning techniques
3	**Open-source software support** Managing and supporting open-source software development projects: cataloging, documentation, review for long-term sustainability, and user support
4	**Cross-mission science and modeling** Addressing cross-mission and division science as well as dedicated support for modeling communities to better integrate and fuse observational data into models
5	**User support** First and second tier of **support for user community** Capable of answering scientific questions about data and information, evaluating and merging community contributions to open-source software, addressing technical questions, and developing unique software to serve their science communities.


Multi-Mission Algorithm and Analysis Platform (MAAP)
Science-focused, cloud-based environment to discover, process, analyze, and share NASA and ESA data
Interactive Storytelling Platform
Interactive exploration, kick-started by NASA's COVID-19 dashboard activities
Common Metadata Repository
Open source, cloud-native, super fast Earth Science catalog and discovery
Earthdata Publication Tool
Centralized authoring and management of NASA Earth Science new products
Distribution and Analysis Platform(s)
Algorithm development and data production
Models and Model Processing
Multi-mission Data Lake
Cumulus
Open source, cloud-native, reusable Ingest and Archive Workflow System
Policies
Open source, open data, articles
Algorithm Publication Tool
Centralized authoring and publishing and discovery tool for NASA Earth Science ATBDs
Advanced Metrics Collection
Cloud-based configurable ingest and archive metrics tracking
NGAP Security, backup, cost controls, scalability
Operational
Active Development
Formulation

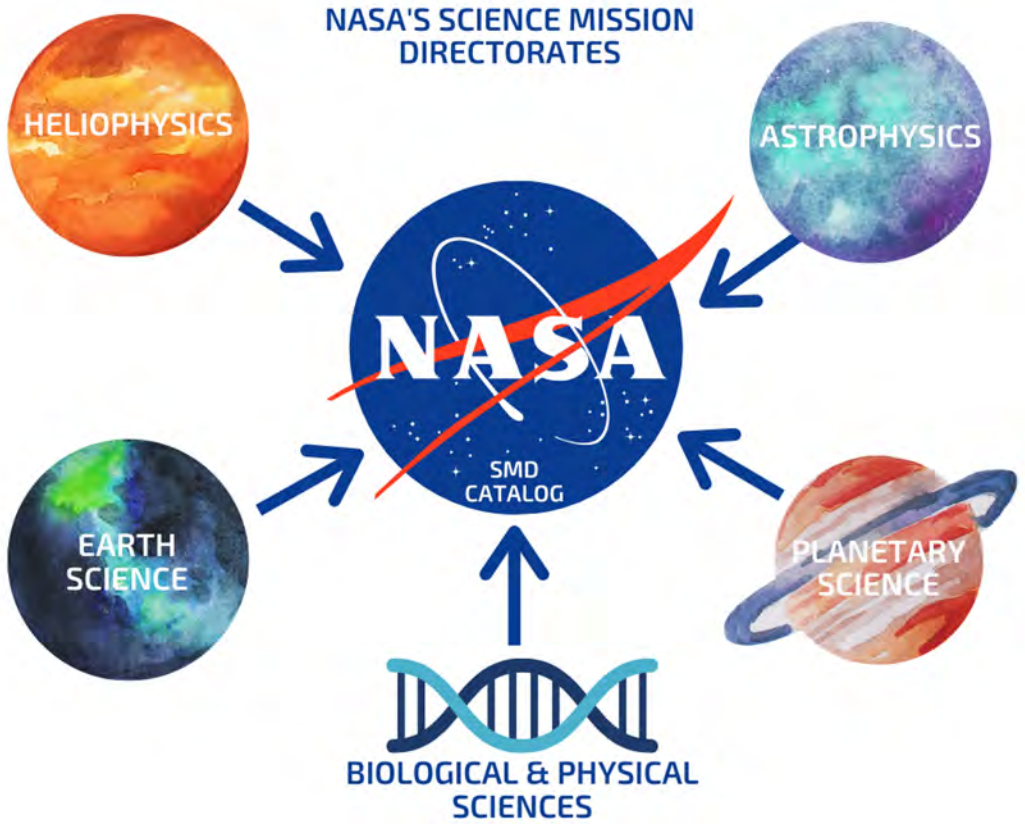
Back up Slides for Right now

NASA's Science Mission Directorate (SMD) Catalog Project

Kaylin Bugbee
Marshall Space Flight Center
October 14, 2021

SMD Data Catalog Goals

- Develop and implement an SMD data catalog to support discovery and access to complex scientific data across Divisions.
- Create a community of practice around data and information standards.



Motivation: Enable collaborative interdisciplinary science

"When we try to pick out anything by itself, we find it hitched to everything else in the Universe."

John Muir

Image Credit: NASA/CXC/JPL-Caltech/STScI

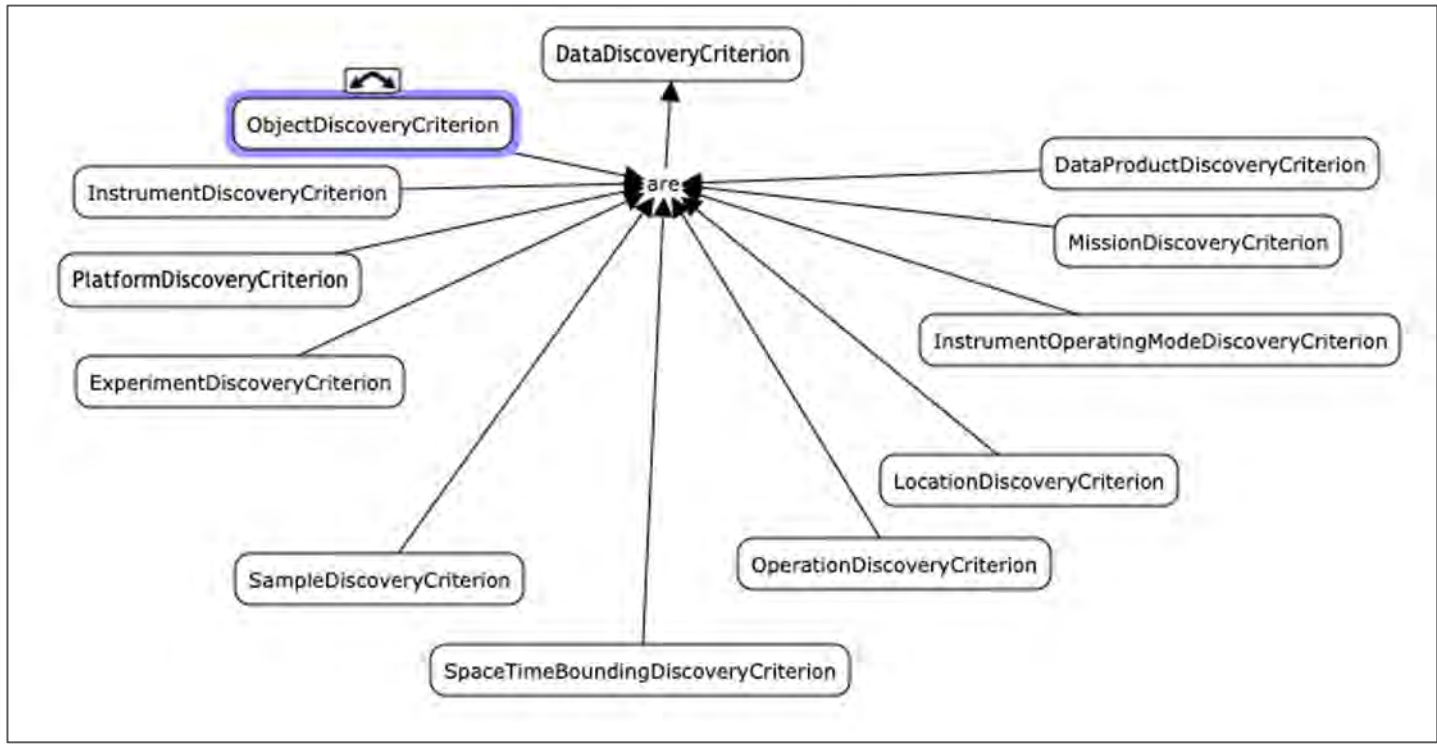
Motivation: Enable Open Source Science

- Data is a key component in making science more open, transparent and reproducible
- Open data and information makes interdisciplinary and collaborative science possible
- Our data is a significant public investment that is also a trusted source of scientific data

Figure Source: Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From open data to open science. Earth and Space Science, 8, e2020EA001562. https://doi.org/10.1029/2020EA001562

SMD Data Catalog: Accomplishments To Date

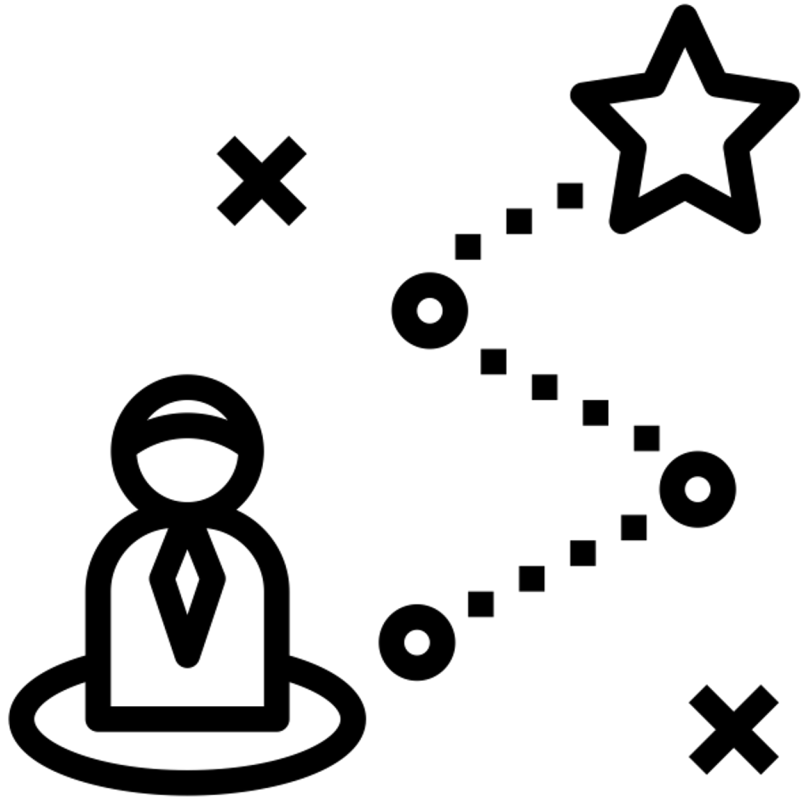
- Established a collaborative working group
- Collected 16 interdisciplinary use cases
- Conducted two deep-dive use case workshops
- Defined the common data discovery concepts across the 5 divisions
- Ran a series of prototyping efforts including a collaborative proof-of-concept with the Enterprise Data Platform Development team



A subset of the SMD data discovery conceptual model.

SMD Data Catalog: Future Plans

- Roll out version 0 of the catalog over the next fiscal year.
- Develop a number of capabilities to support interoperability including an SMD wide vocabulary and schema matching across data models.
- Continue prototyping efforts to augment search capabilities.



Created by Nithinan Tatah
from Noun Project

Project Team and Collaborations

SMD Catalog Team

- Kaylin Bugbee (NASA MSFC)
- Christian Reyes (Project support)
- Shawn Foley (MITS II)
- Mark Parsons (UAH)
- Ruth Duerr (Ronin Institute)
- Ashish Acharya (UAH)
- Ahmed Eleish (RPI)
- Emily Foshee (UAH)
- David Bitner (DevelopmentSeed)
- John Hedman (DevelopmentSeed)
- Tammo Feldman (DevelopmentSeed)
- Necoline Hubner (DevelopmentSeed)
- Leihla Pinto (DevelopmentSeed)
- Dai-Hai Ton That (UAH)
- Dan Berrios (USRA/Ames)
- Rick Plotka (RPI)
- Peter Fox (RPI)

Enterprise Data Platform

- Charles Driessnack (MSFC)
- Wes Adams (MSFC)
- Tobie Smith (MSFC)
- Stephen Tyszka (MSFC)
- Jason Duley (HQ)
- John Monroche (MSFC)
- Brandon Shaffer (MSFC)
- Melania Stewart (MSFC)

Working Group

- Alan Smale (Astrophysics)
- David Ciardi (Astrophysics)
- Graham Berriman (Astrophysics)
- Robert Candey (Heliophysics)
- Brian Thomas (Heliophysics)
- Dan Berrios (Biological & Physical Sciences)
- Emily Law (Planetary Science)
- Thomas Morgan (Planetary Science)
- Steve Hughes (Planetary Science)
- Erich Reiter (Earth Science)

Interested in collaborating or learning more? Please direct questions to:
Kaylin.m.bugbee@nasa.gov
christian.g.reyes@nasa.gov

AI/ML

Science Mission Directorate
Earth Science Data Systems

OSSI SMD AI activities

Manil Maskey, Ph.D.

OSSI Workshop
October 14, 2021

OSSI SMD AI team

Srija Chakraborty
AI SME/Post Doc

Dan Duffy
HECC

Lika Guhathakurta
Heliophysics

Mike Little
HECC

Megan Ansdell
Planetary Science

Roopesh Ojha
Astrophysics

Yvonne Ivey
SDMWG

Sylvain Costes
Biological and
Physical Science

Evan Scannapieco
Astrophysics

Manil Maskey
Earth Science/Lead

NASA Science Mission Directorate (SMD) Artificial Intelligence and Machine Learning (AI/ML) Initiatives

SMD's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024 Report identified that AI/ML has yet to be fully appreciated and understood by SMD and science disciplines

Strategy 2.4: Invest in the tools and training necessary to enable breakthrough science through application of AI/ML

Recommendation 11: SMD should make investments to incentivize and educate the community on how to use AI/ML to approach science in new ways. Hands-on training can be achieved through expansion of hackathons, competitions, and grant programs. Science results and lessons learned about the use of AI/ML will be shared at community meetings to increase awareness of the potential of these techniques.

Activities:

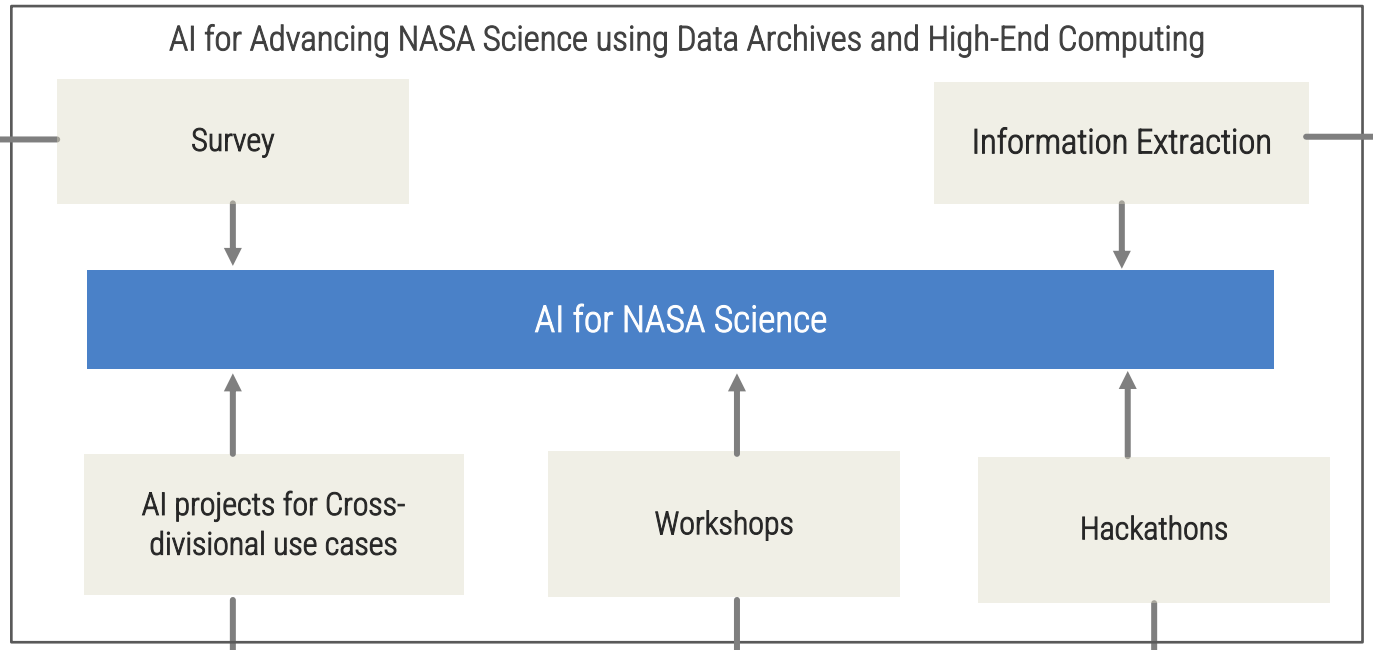
- Identify areas of natural collaborations on AI/ML across SMD
- Conduct expert workshop on AI for science
- Explore industry partnership
- Develop a roadmap to leverage **large volumes of data and computation** to accelerate AI/ML across NASA science

SMD SDMWG AI/ML

Goal of determining the Current state of the art, Workforce, Policies, Datasets, Computation, Needs for AI/ML

Participants: NASA and NASA funded Researchers

Goal of building a knowledgebase of AI activities within NASA science objectively; understanding trends in AI adoption, challenges, datasets, recommendations using scientific literature, proposals, projects



Goal of demonstrating early win projects in utilizing AI for cross divisional use cases within NASA science, understanding challenges,

Participants: NASA funded researchers

Goal of addressing challenges in AI for NASA science from data and computation perspective, discussing tools and framework, strategy for making NASA data AI-ready, benchmarking, and open sharing for competitions and re-use, path forward for accelerating AI research and applications

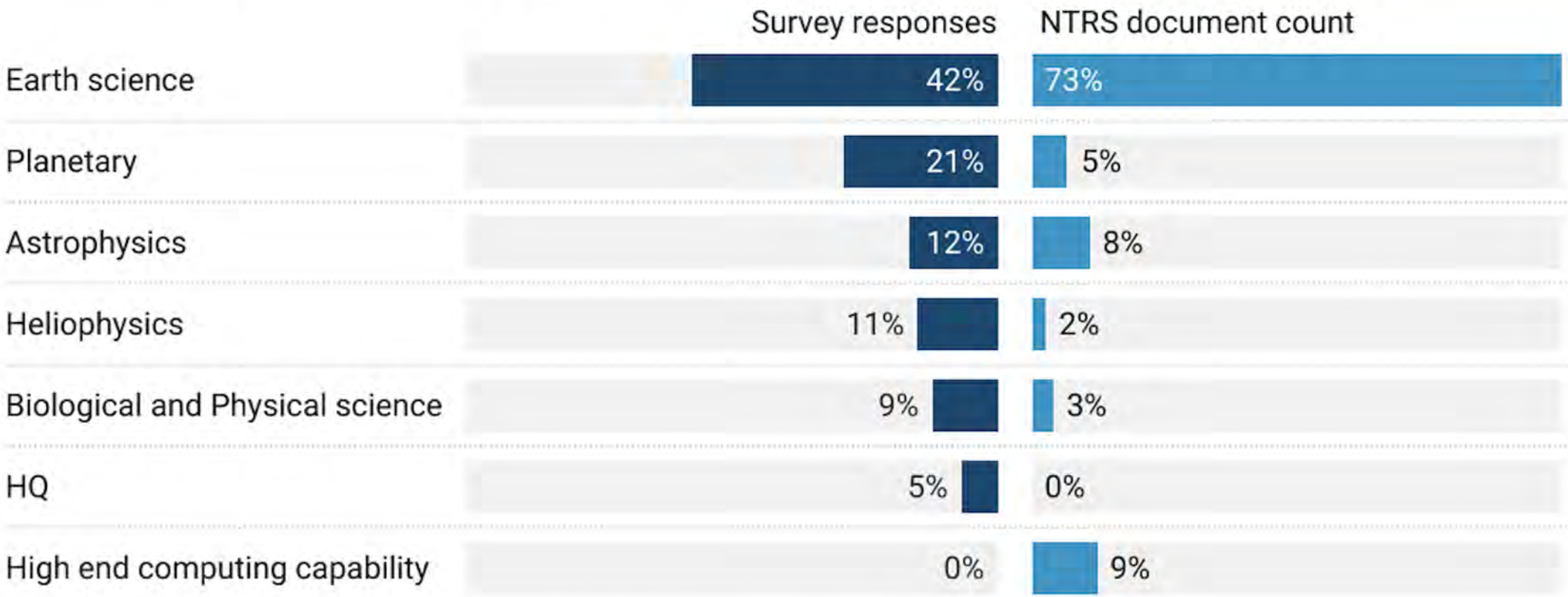
Participants: Industry, Academia and NASA AI experts

Goal of providing AI implementation experience, fostering collaboration – Team Science, utilizing existing benchmark NASA science datasets, utilize computational resources, solve cross-divisional problems, identifying challenges

Participants: Anyone (in US)

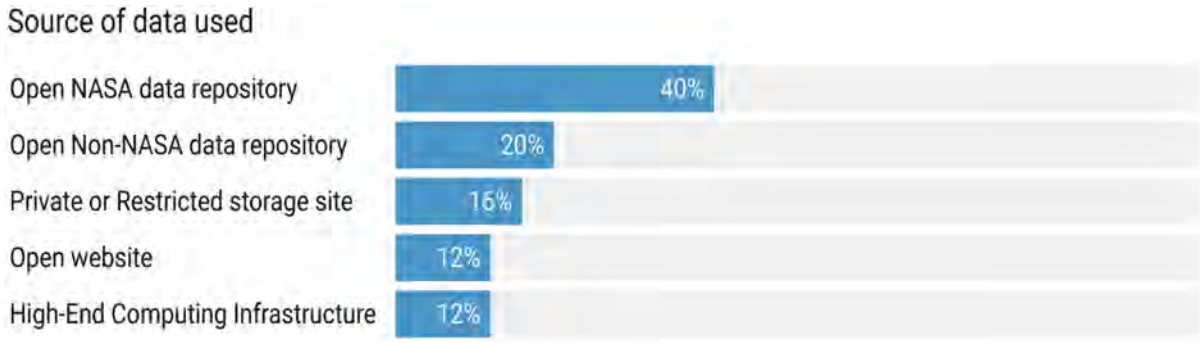
AI within NASA SMD Divisions/Programs

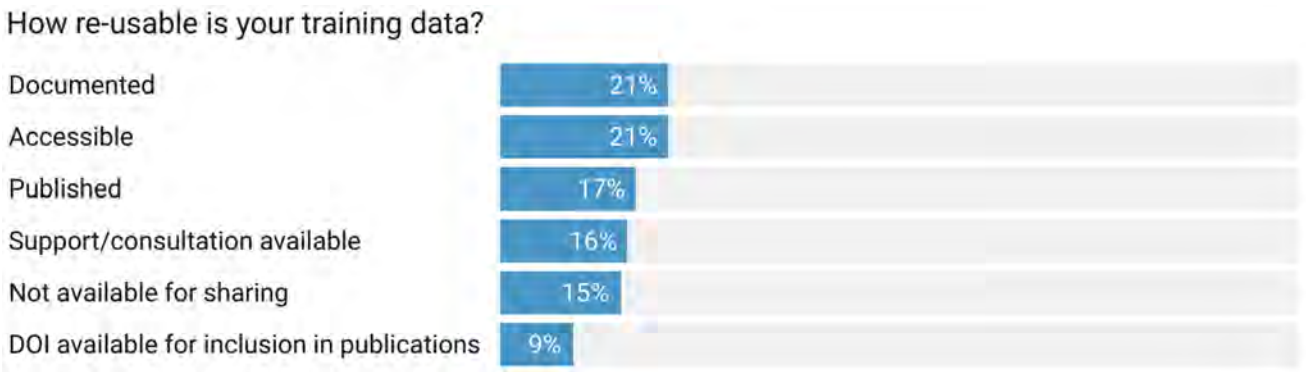
SMD AI Survey Responses and NTRS Documents

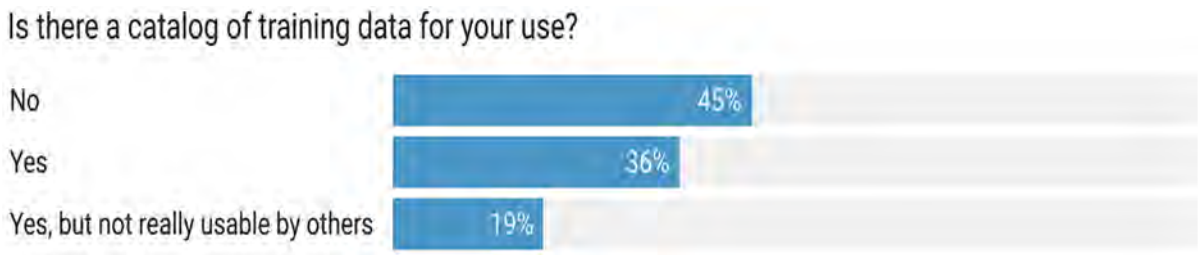


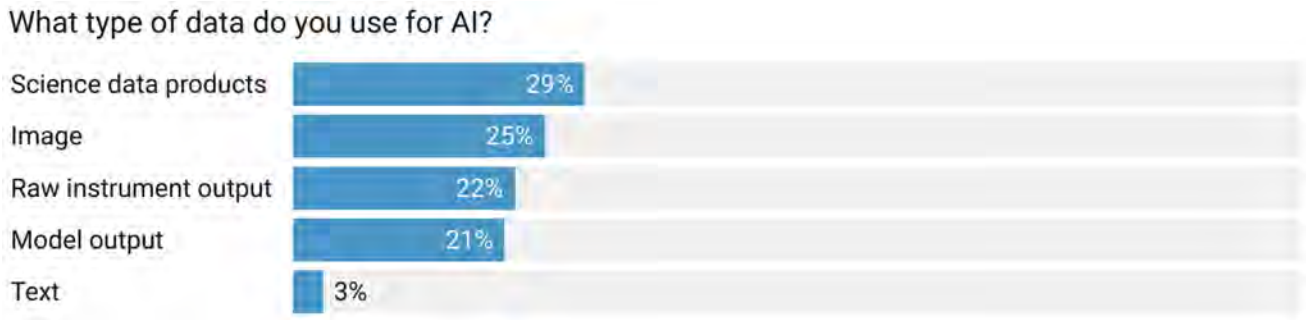
560 total survey responses 8317 total NTRS documents

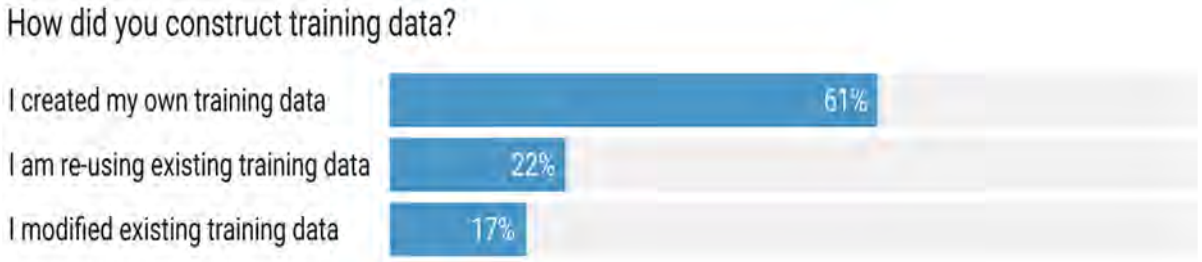
Survey response - AI and data

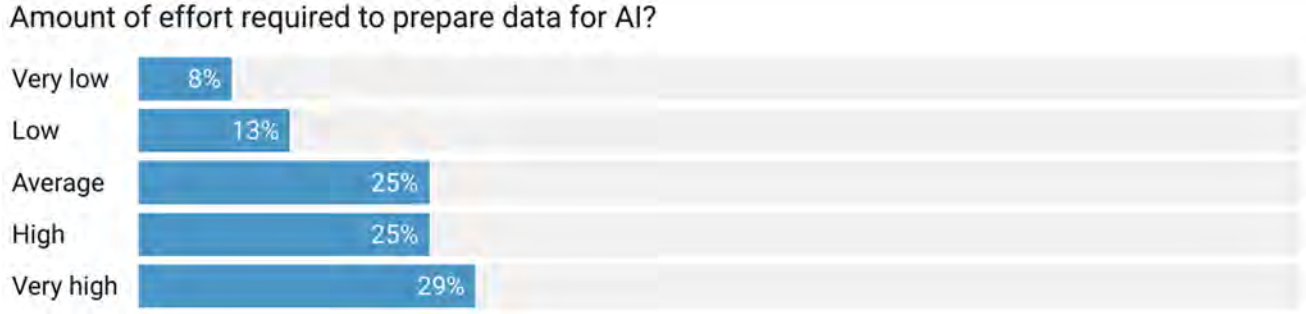




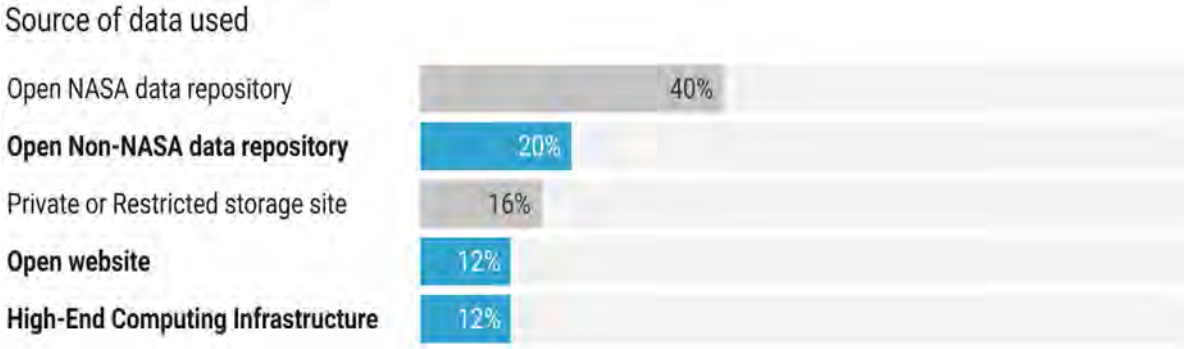


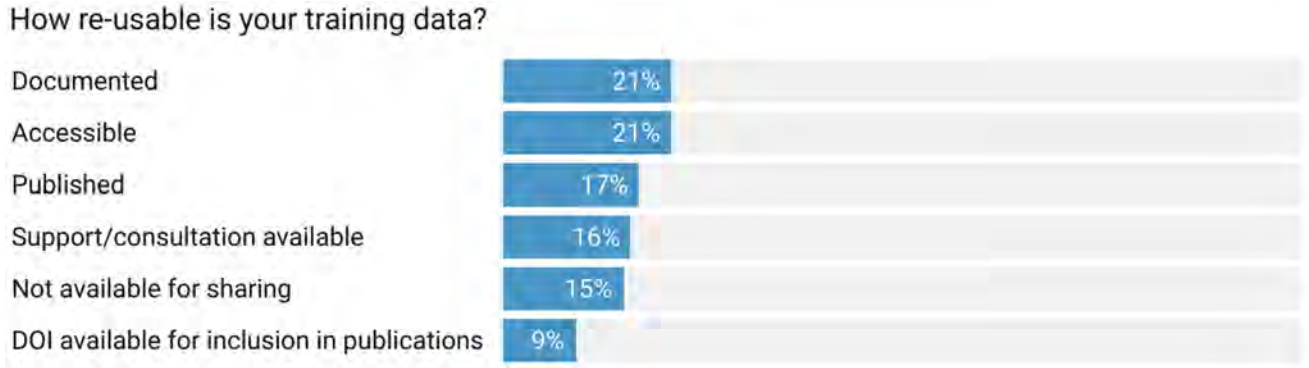


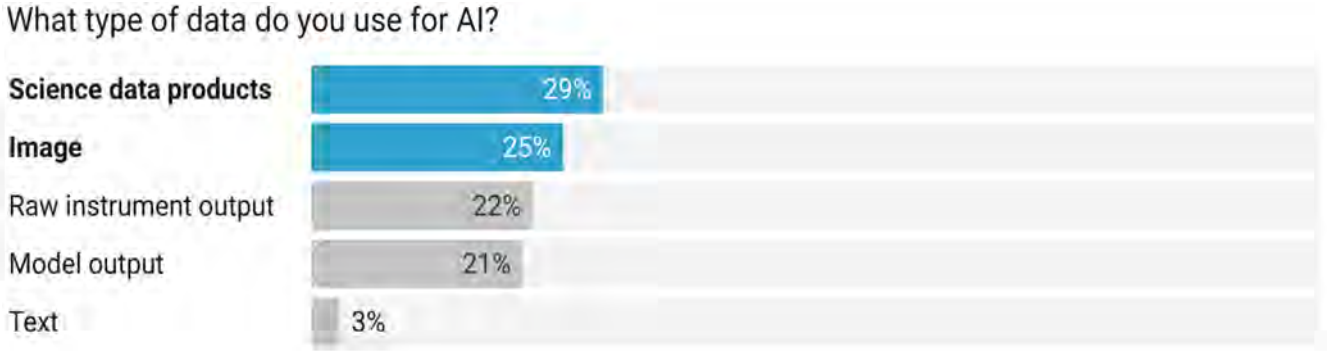


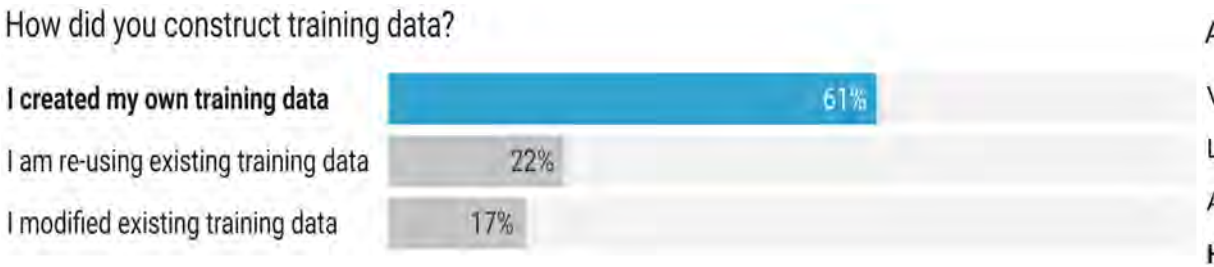


Survey response - AI and data









Cross-divisional use cases - Prototype AI projects

Domain-Agnostic Outlier Detection in Science Datasets - *Earth science + Astrophysics + Planetary*

Leveraging AI to Perform Pixel-Level Extraction, Classification, and

Segmentation of Astrophysics and Earth Science Imaging Data - *Earth science + Astrophysics*

Petabyte scale search on multi-spectral unlabeled data to rapidly curate annotated

datasets + Search By Image for NASA Science - *Earth science + Astrophysics*

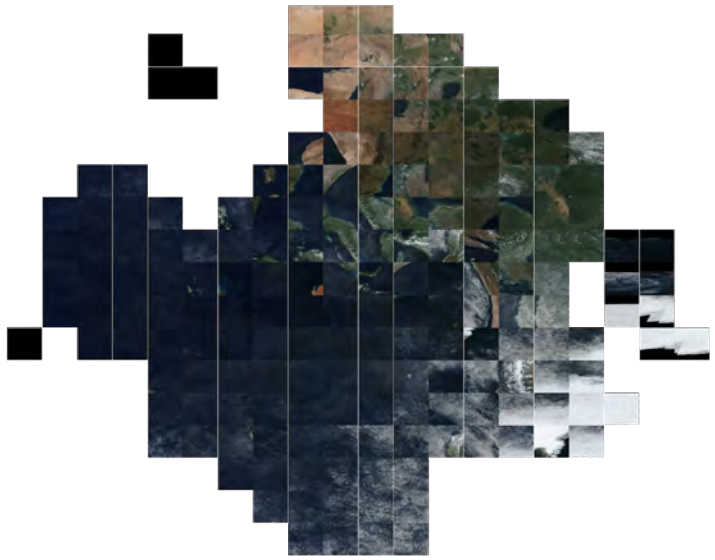
Enhancing NASA's Science using Physics Informed Deep Learning - *Heliophysics + Planetary*

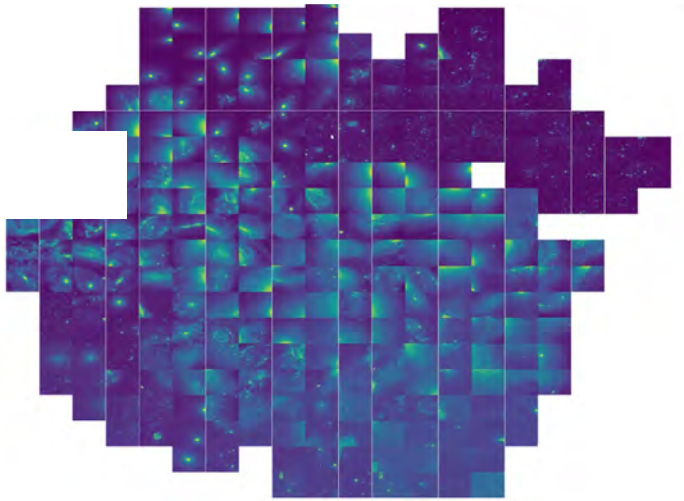
A Bayesian View of the Solar Wind Impact on Mars' Magnetic Environment - *Heliophysics + Planetary*

Cross-divisional use cases - Prototype AI projects

Petabyte scale search on multi-spectral unlabeled data to rapidly curate annotated datasets + Search By Image for NASA Science - *Earth science + Astrophysics*

PI: A. Koul & J. Peek



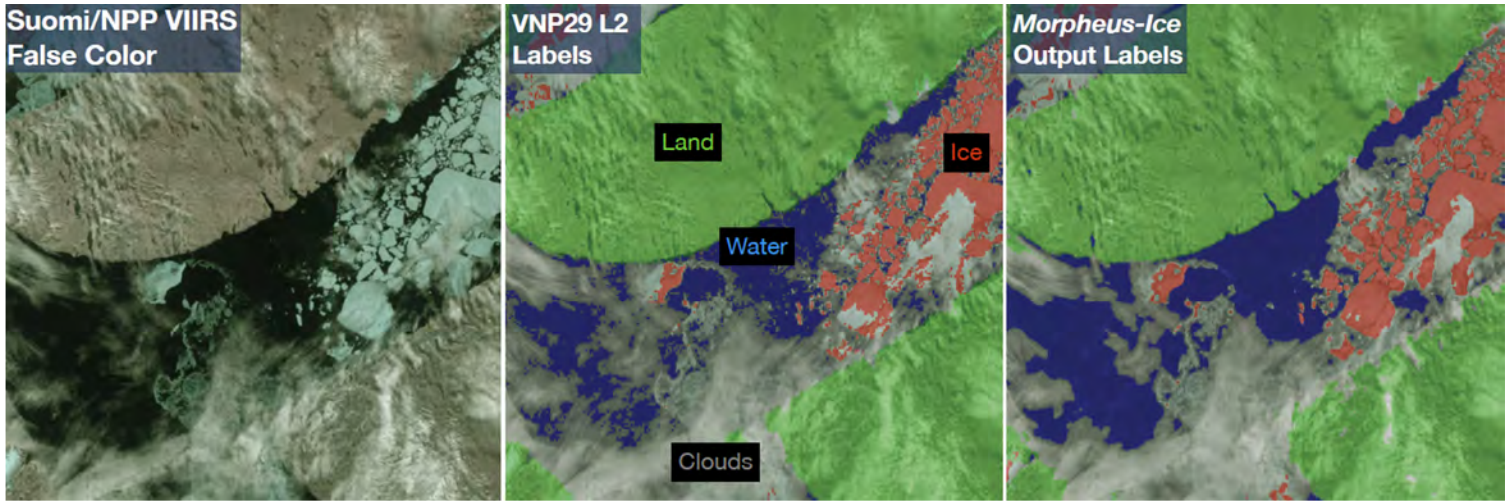


Cross divisional AI project: Self supervised learning approach developed for GIBS archives applied to Hubble telescope data

Cross-divisional use cases - Prototype AI projects

Morpheus-Ice: Pixel-Level Classification of Imaging Data from Astrophysics to Earth Science

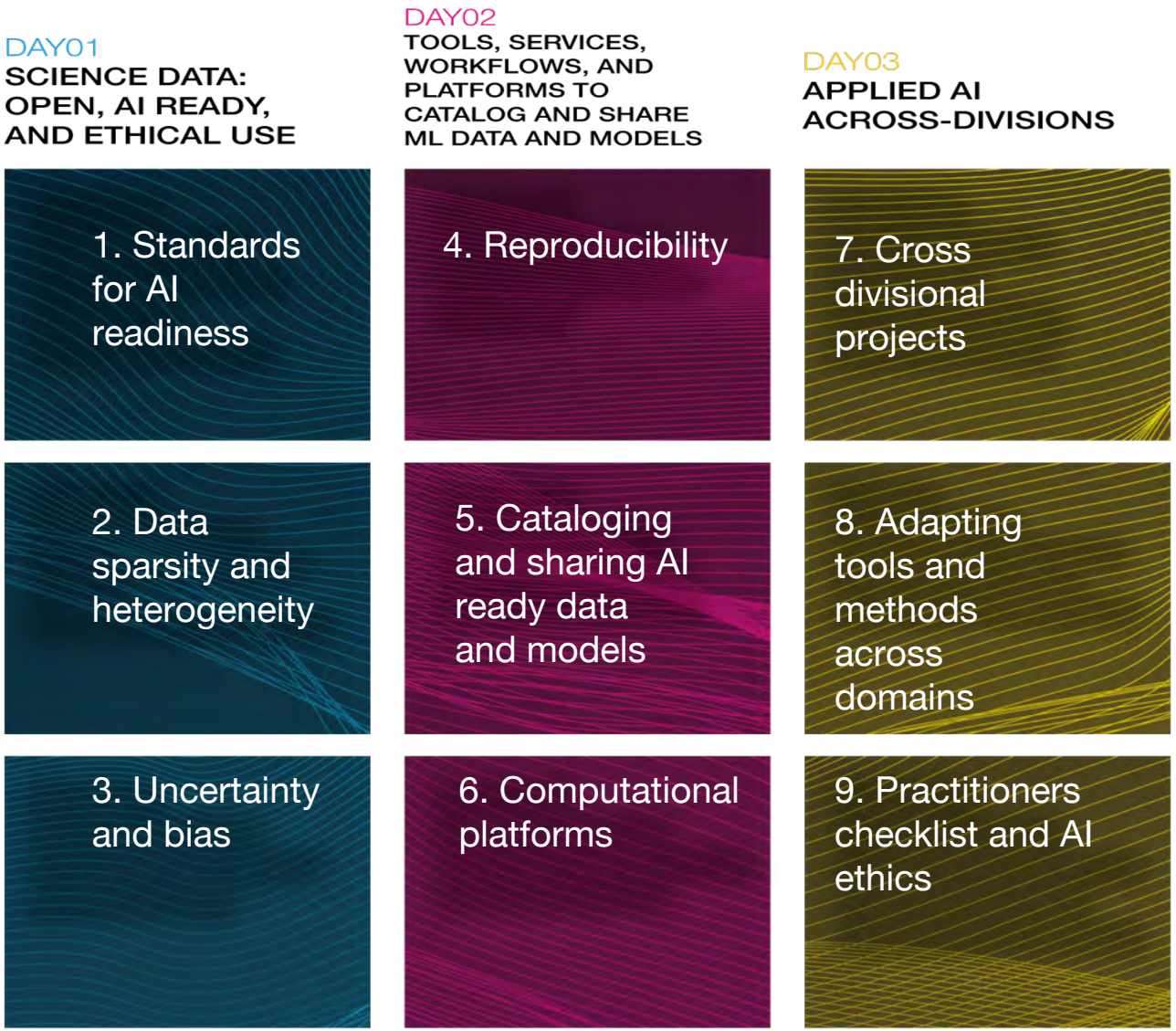
PI: B. Robertson



Level 2 data products (VNP29) available from NSIDC that label land, water, clouds, sea ice, night, and bad pixels, created by identifying the relative reflectance of these features at different wavelengths.

Directly supply VIIRS reflectance images and labels to Morpheus, retrain with no architecture changes

SMD AI Workshop



Workshop main findings

Development of SME informed AI-ready data standard

Development of reusable AI-relevant data management tools

Development and publication of labelled training data and models for AI applications and benchmarking

Lowering the barrier to entry to access computing resources for AI

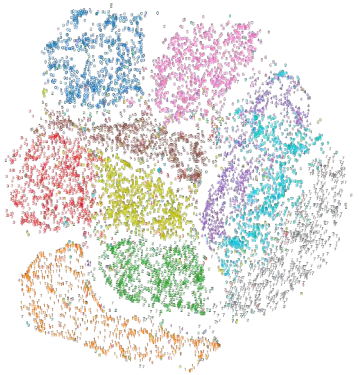
Supporting cross domain collaborations and sharing

Training and education for AI skill development

Incentives for reproducibility and open sharing of AI artifacts

Embedding of ethical considerations of AI into science research processes



Thank you.

Prizes and Challenges

Cross-Divisional Prizes and Challenges

Katie Baynes
kathleen.baynes@nasa.gov

Cross-SMD Prizes and Challenges

This effort represents an early open source science inspired initiative to use prizes and challenges to help support the overall strategy and spur innovative, cross-divisional ideas, engaging both internal NASA community and eventually the general public.

Chair: Katie Baynes (ESD)

Team: Quang-Viet Nguyen (JASD), Steve Crawford (SMD)

CoE Collaborative Innovation: Carissa Callini, Steven Rader

Building off Past Successes: CoECI and the NASA Tournament Lab

The Center of Excellence for Collaborative Innovation (CoECI) collaborates with innovators across NASA and the Federal Government to generate ideas and solve important problems by working with global communities via the NASA Tournament Lab.

https://www.nasa.gov/coeci/ntl

NASA @ Work Internal Solicitation

Using NASA Science Data and Computing for Cross-Disciplinary Science

Do you have an idea for using data from Earth observation to help us better understand physical interactions on other planets?

Have you thought about how to use data about our sun to inform our understanding of other systems?

:::

NASA's Science Mission Directorate Strategic Data Management Working Group (SDMWG) is looking for ideas that could be turned into topics for significant prizes and challenges focusing on utilizing NASA's free and open science data from multiple science disciplines. Science Mission Directorate (SMD) has allocated to put toward these challenge ideas. We expect up to 7 challenge ideas to be selected for funding and implementation.

Browse Ideas Subscribe to Challenge

Multi-Staged Solicitation and Execution

NASA@Work Phase 1

Broad audience, broad ideas, large number of submissions expected, brief summary of problem/challenge

NASA@Work Phase 2

Narrowed field of candidates (10 max), focused workshops on refining solicitation for public release

Internal Preparation

Establish and coordinate relationships between science divisions and selected candidate activities, preparing for public solicitations (2 projects selected)

Execution Phase

Continued partnering with CoECI (potentially other groups as appropriate) to conduct selected prizes/challenges within targeted public communities

October 2020 — Spring 2022

Leveraging ML for Data Constrained Planetary Mission Instrumentation

PI: Victoria da Poian, GSFC

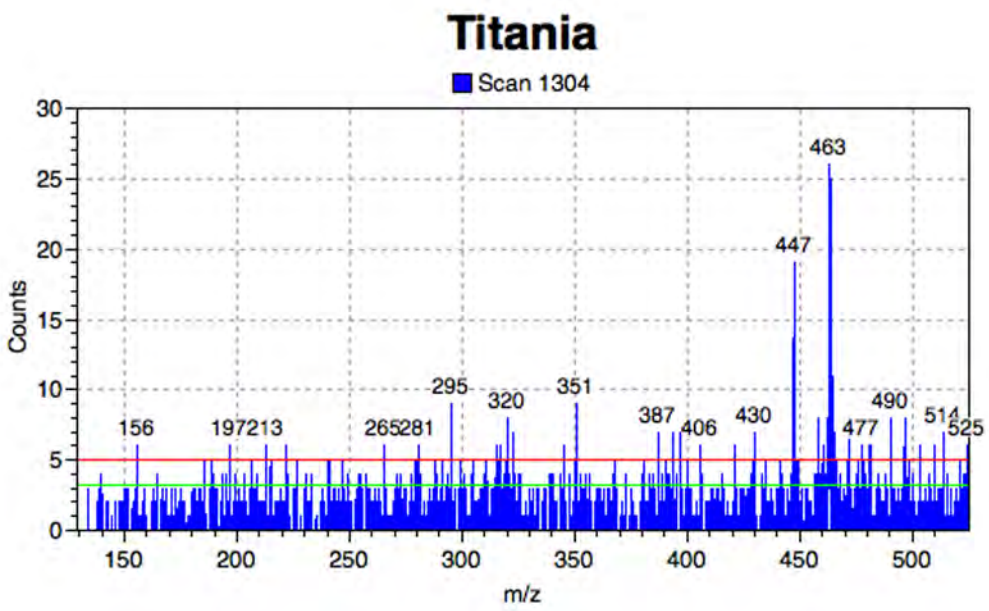
Having a clear a streamlined pipeline supporting planetary science exploration missions

Short-term: Earth-based in order to support scientists in their decision-making process

Long-term: on the instrument, onboard the mission to enable science autonomy

Challenge Objective:

To develop innovative approaches that can most accurately classify the chemical composition of material samples taken on planetary missions by leveraging data collection terrestrially



DRIVENDATA

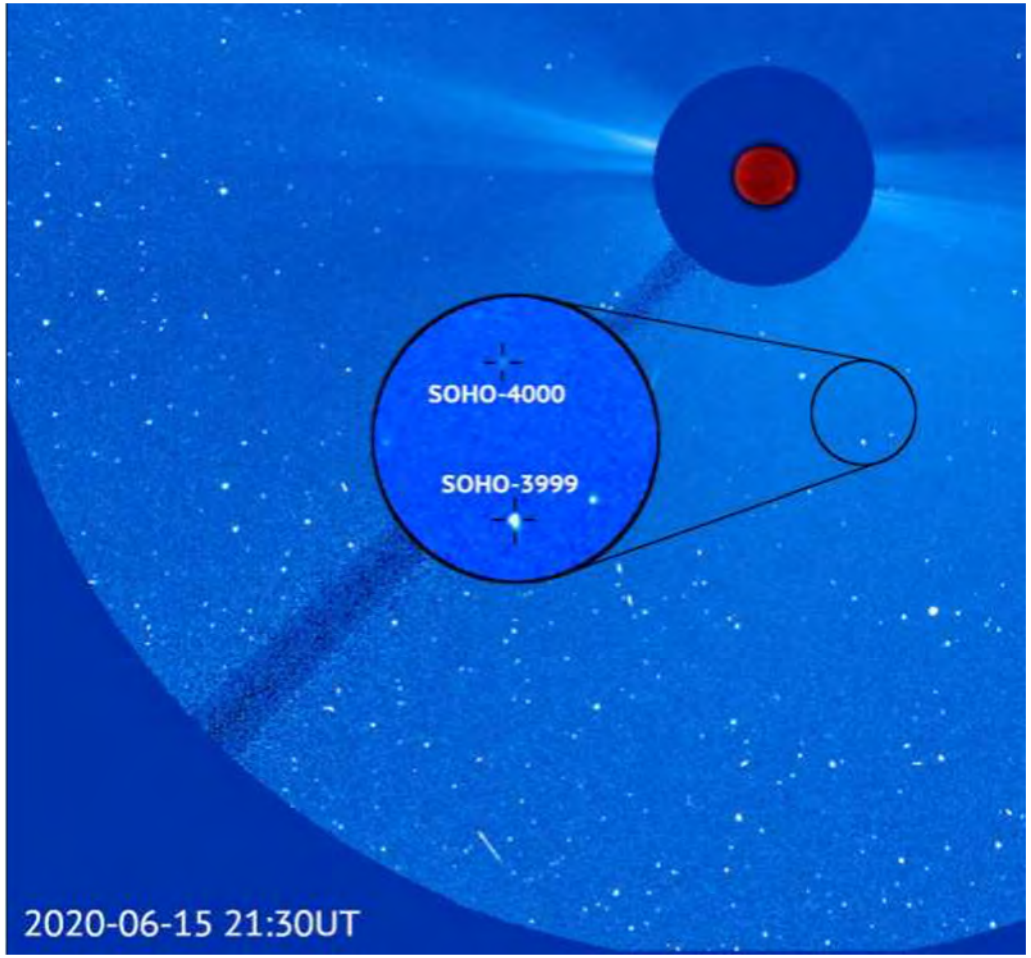
Comet Detection using SOHO/LASCO data

PI: Ekaterina Verner, HQ

During more than 25 years of operation SOHO/LASCO instrument identified more than 4200 comets, most of them being very faint.

Challenge Objective:

To build AI/ML based approach which can help to identify the faintest noise-level comets by reducing the background noise in the SOHO/ LASCO data.



topcoder™

Citizen Science

EXPLORE SCIENCE

Citizen Scientists/Open Source Scientists

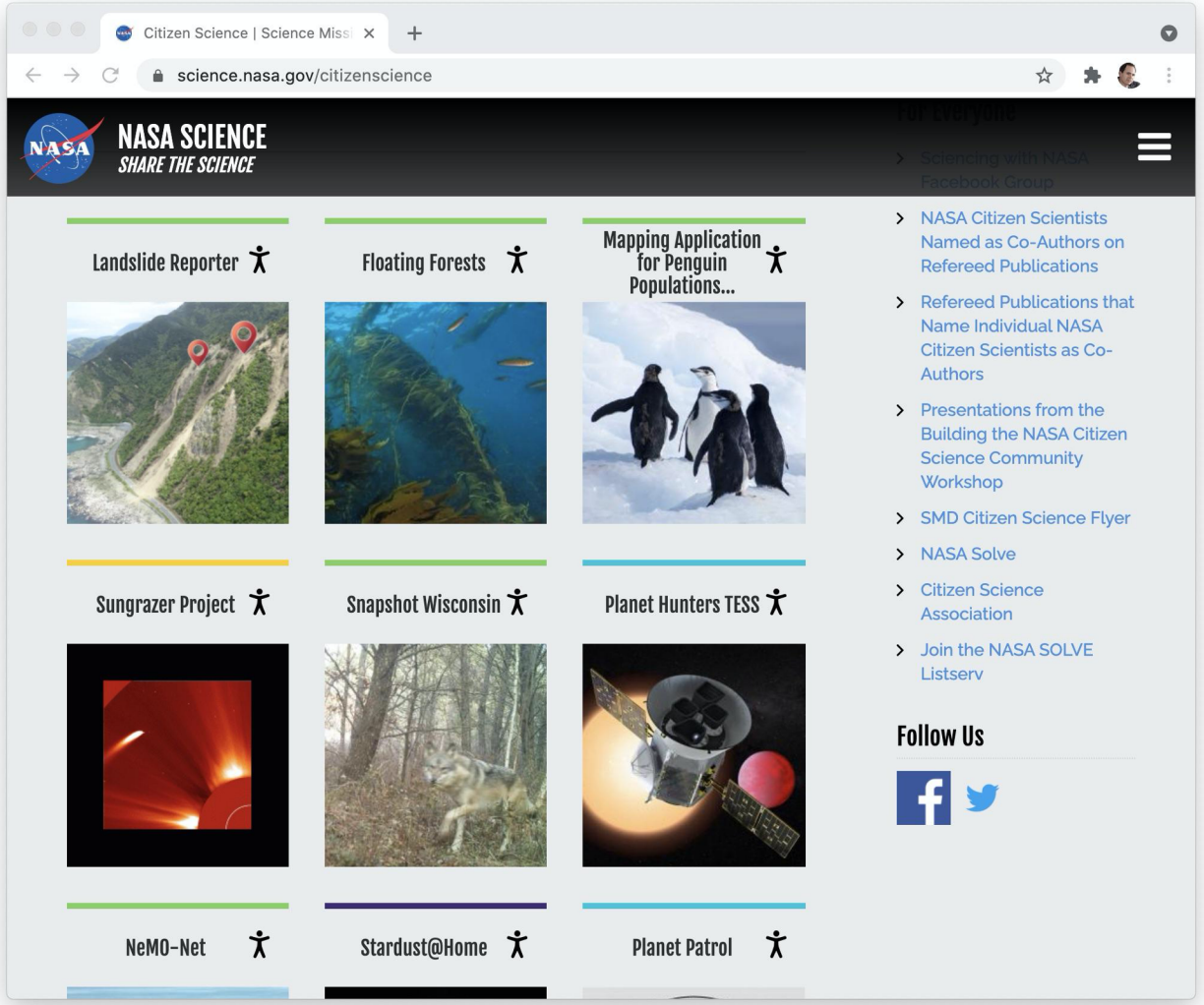
Marc Kuchner, Citizen Science Officer, SMD

October 13, 2021

NASA’s citizen science projects are **science projects that rely on volunteers.**

science.nasa.gov/citizenscience

25 active NASA projects online



NASA's citizen science projects now dominate their scientific fields. They have discovered:

- **Most** of the known comets
- **All** of the known samples of interstellar material
- **Half** of the ultracool brown dwarfs
- **Most** of the long period (>2 yr) extrasolar planets

NASA citizen scientists have discovered:

- The first extreme T subdwarfs
- Zika virus in Peruvian cemetery vases
- The oldest white dwarf debris disk
- The “Dipper” star phenomenon
- The “Peter Pan” disk phenomenon
- The star-forming regions called “yellowballs”.
- 400,000 Martian seasonal fans
- 283,000 emperor penguin nests
- 8,900 mosquito breeding sites
- 7 meteorites
- One new *kind* of aurora named STEVE

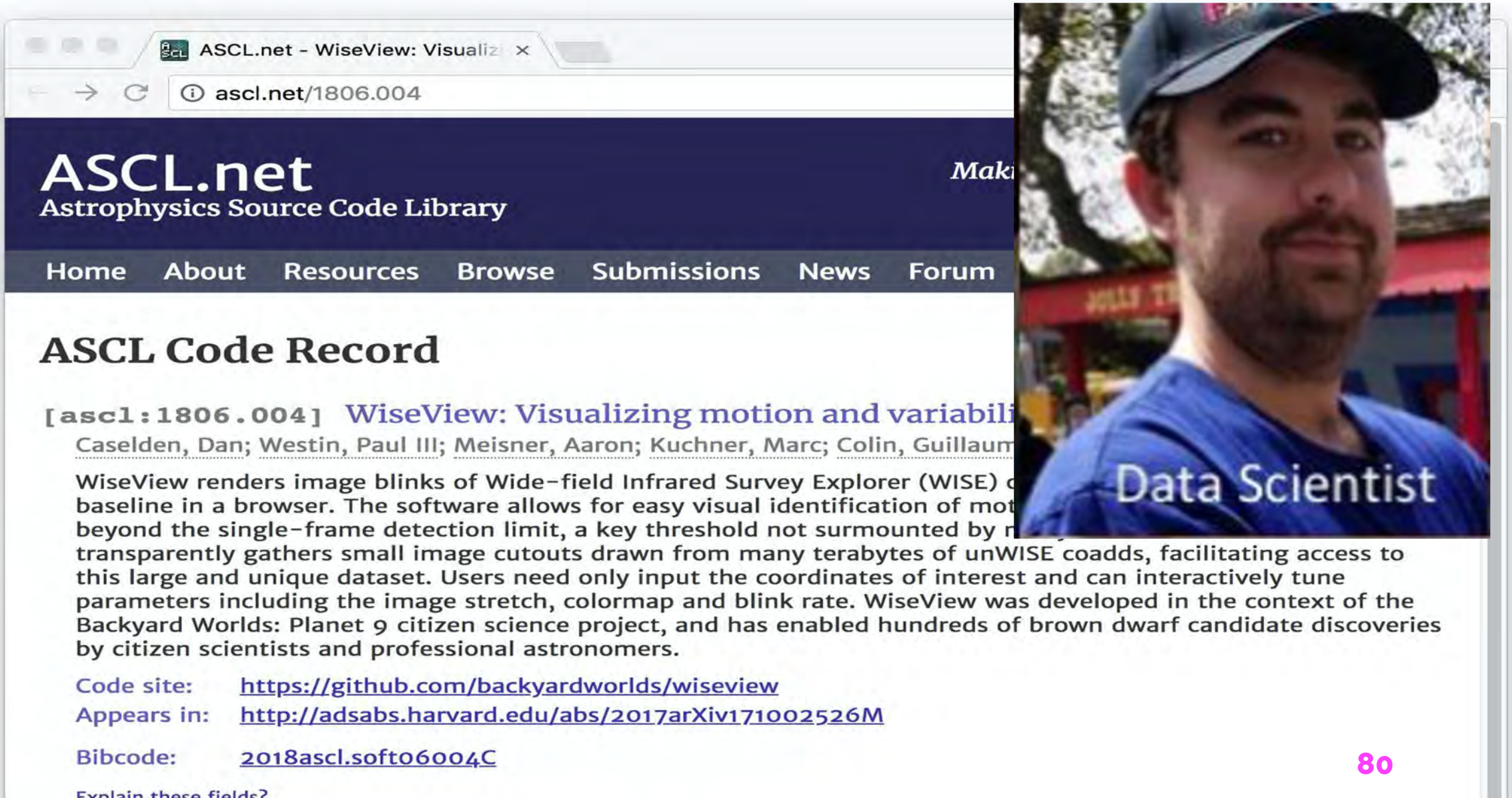
191 NASA Citizen Scientists Have Become Named Co-Authors on Scientific Papers

NASA citizen science involves > 1.5 million volunteers including

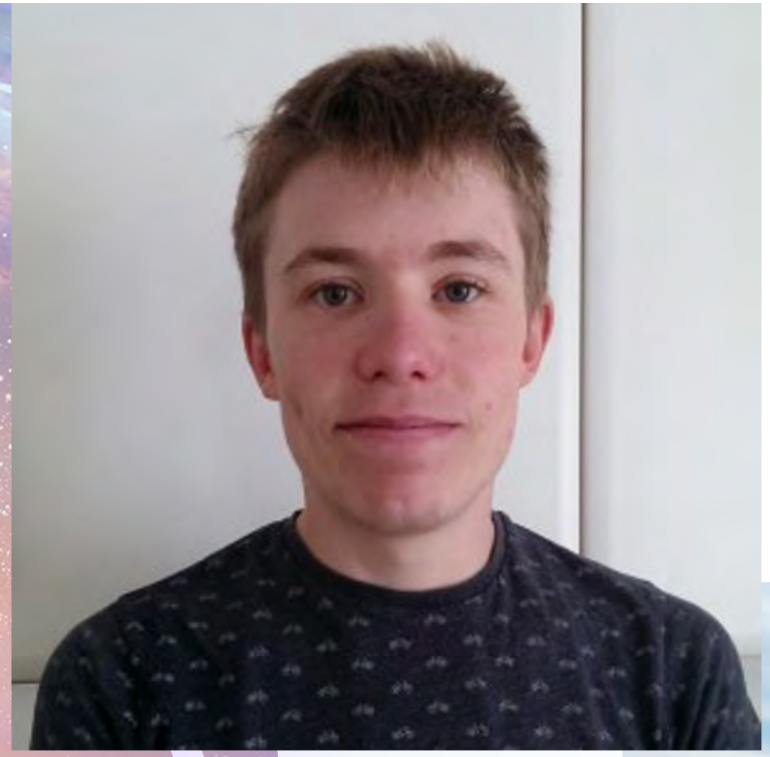
~90,000 people with graduate degrees.

Our Citizen Scientists Write Codes!

Dan Caselden



Our Citizen Scientists Write Codes!



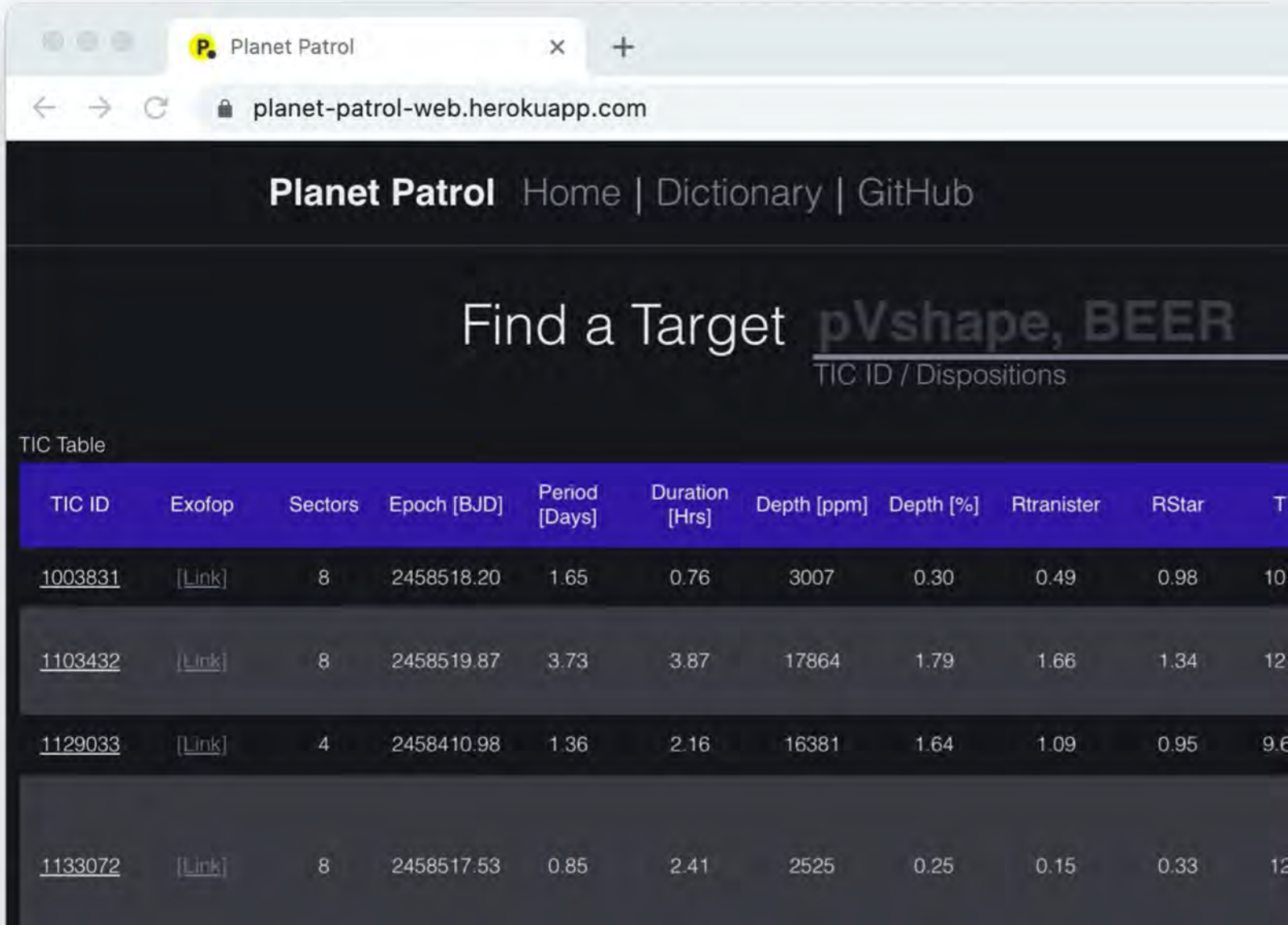
Tom Bolton

Physics Ph.D. student.
His penguin population dynamics model outperformed the model created by ecologists.

Our Citizen Scientists Write Codes!

High School student
Ryan Salik

Wrote vetting tool for
Planet Patrol



However, citizen scientists:

- don’t know our jargon.
- don’t know our culture
- don’t know where to find our data and tools

90,000 NASA citizen scientists with graduate degrees.

Keep them in mind as you work to open science!

Astrophysics Data System

The NASA Astrophysics Data System: a Discovery tool for Open Science

Alberto Accomazzi & the ADS Team
aaccomazzi@cfa.harvard.edu | *@aaccomazzi*
NASA Astrophysics Data System | *@adsabs* | *https://ui.adsabs.harvard.edu*

NASA Open Source Science Workshop - October 14, 2021

Building a Better Information System for Open Science

NASA SMD research requires expertise spanning across boundaries

- Astronomy & Astrophysics
- Planetary Sciences, Geophysics, Astrobiology
- Solar Physics, Space Weather, Plasma Physics

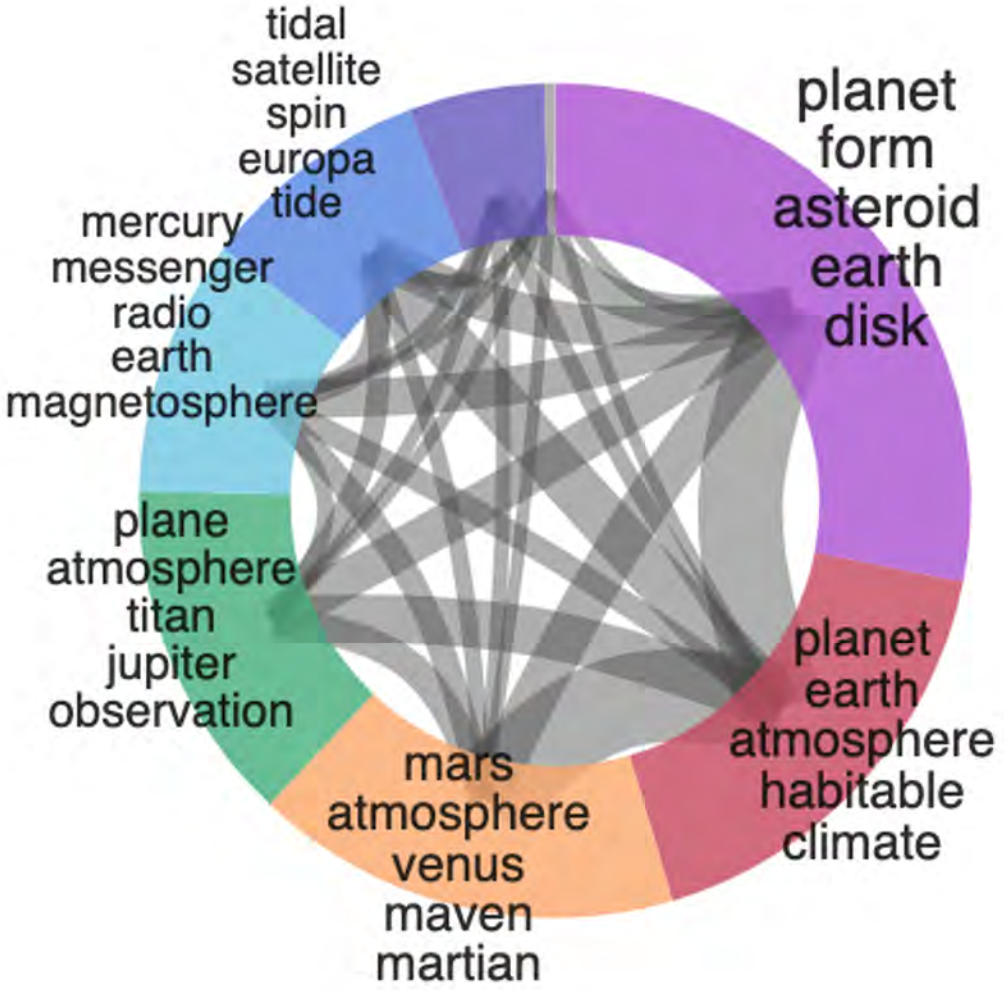
Literature can be seen as central, organizing point to navigate research fields

- Big challenges require communities of experts from different fields working together
- As interdisciplinary research develops, different fields become organically connected and discoverable through topics, citations, co-readership

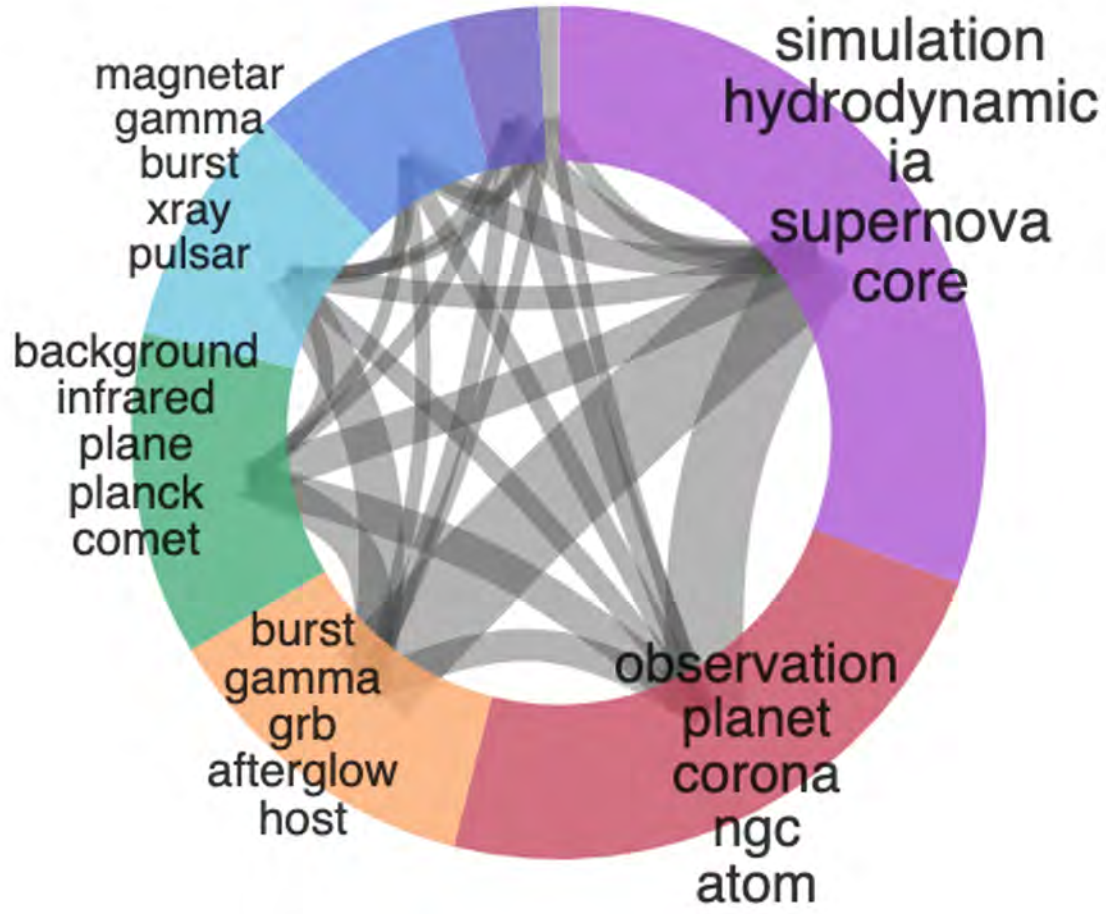
Connections between literature, software and data products increase discovery of all research artifacts

- Links to archives crucial for making data more discoverable and shared

Cross-Disciplinary Analysis (2020)



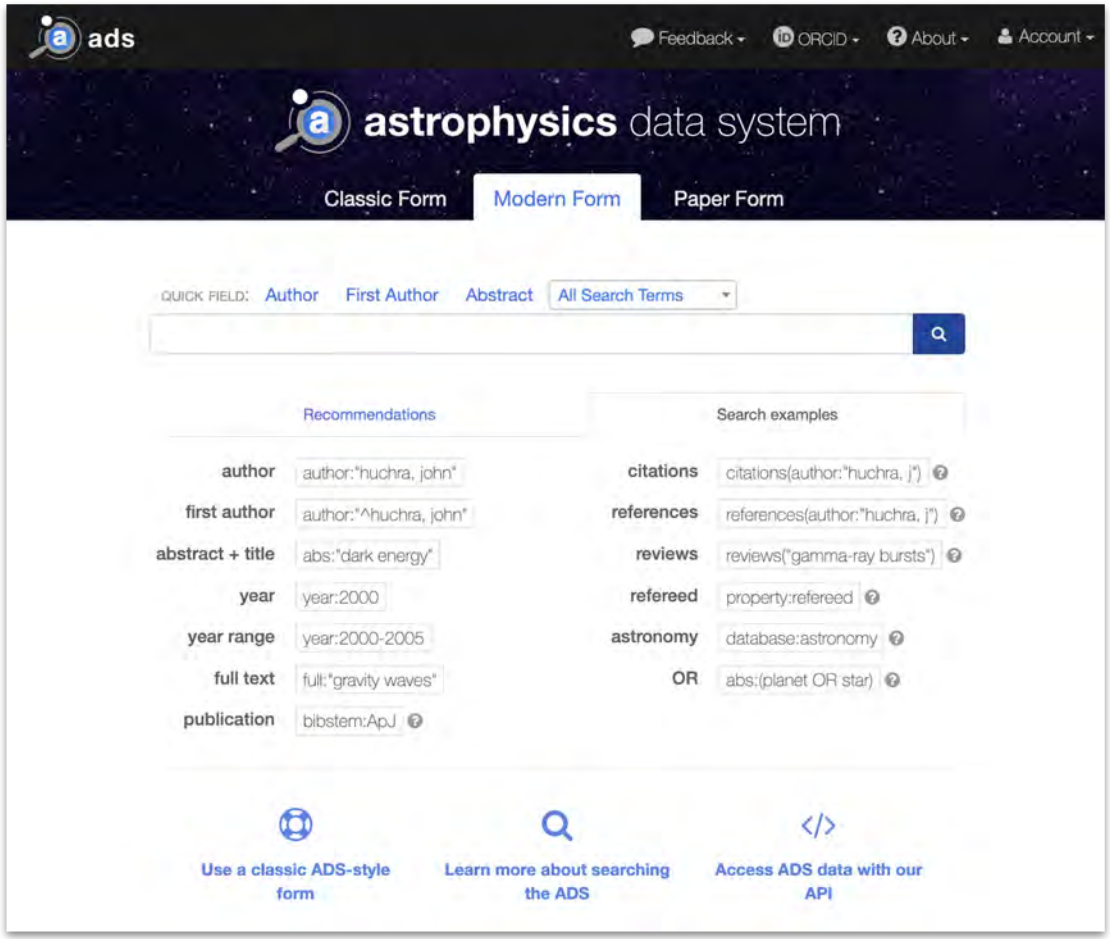
A subject matter clustering of recent cited Planetary Science and Astrobiology literature from the 2020 papers discussing exoplanets. Over 9% of the cited exoplanet literature appears in solar system papers.



A subject matter clustering of recent Astrophysics papers referencing articles funded by the NASA Heliophysics division.

The NASA Astrophysics Data System (ADS)

- ADS is a NASA-funded project which provides discovery services for scholarly literature in Astronomy & Physics
- 15M metadata records, most of them traditional publications
- 6M full-text documents from all major publishers
- A citation graph with over 8M nodes and 142M edges
- (Anonymous) usage data for 50k regular users

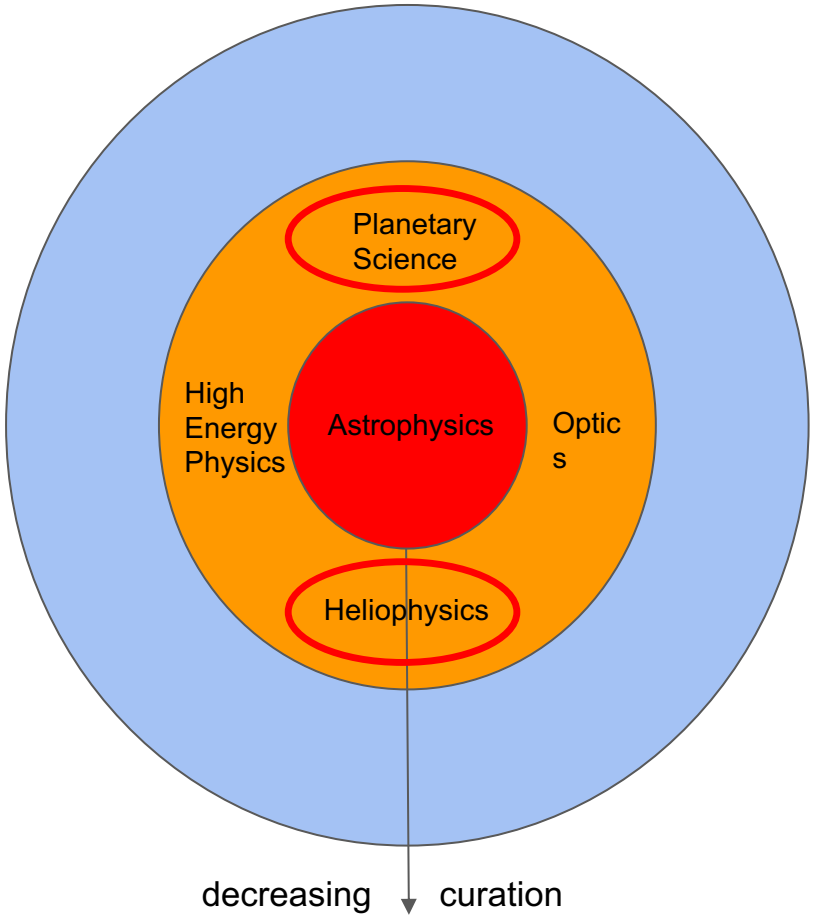


https://ui.adsabs.harvard.edu

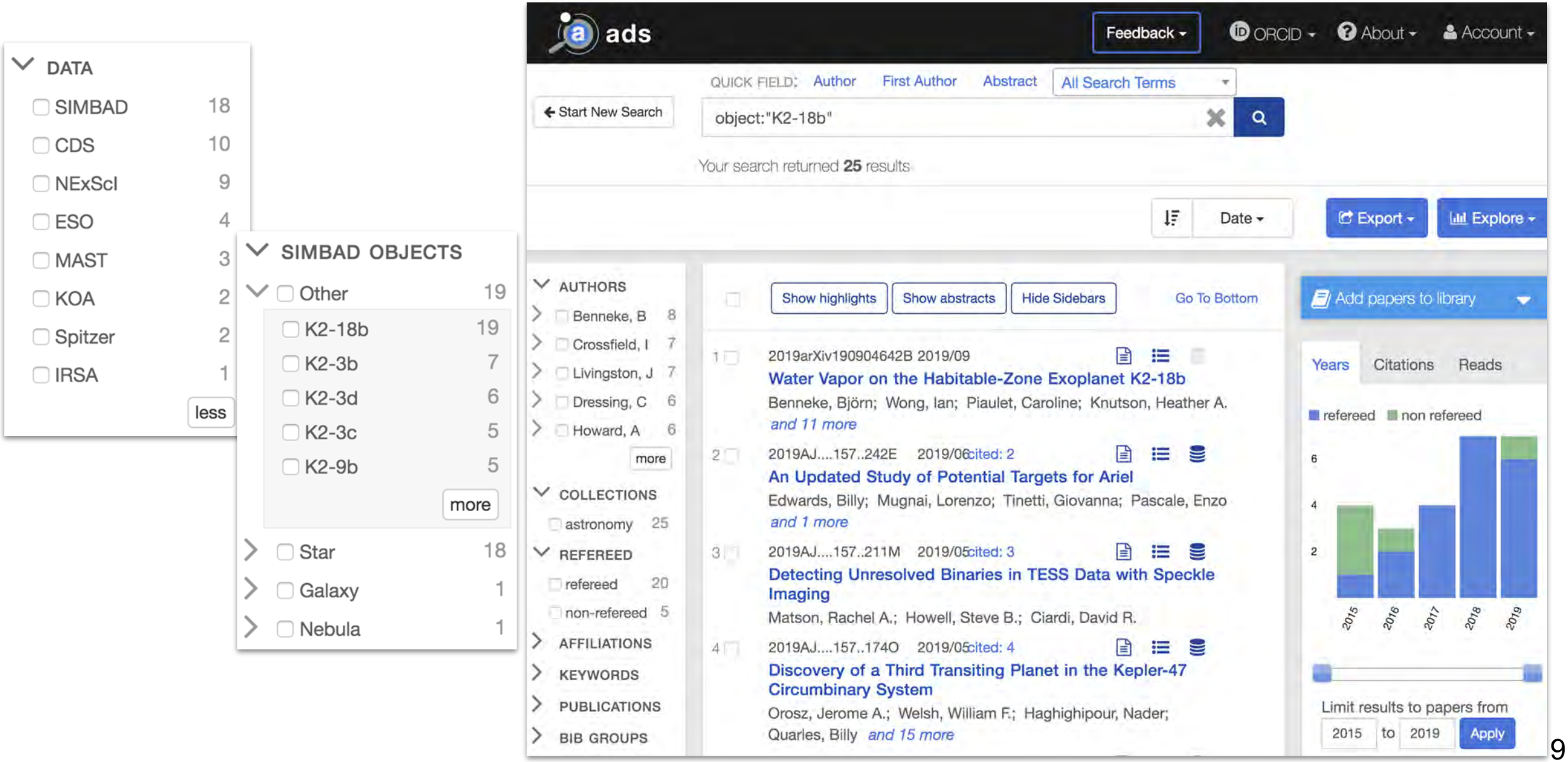
Curation Levels of ADS Content

Core collection: Astrophysics

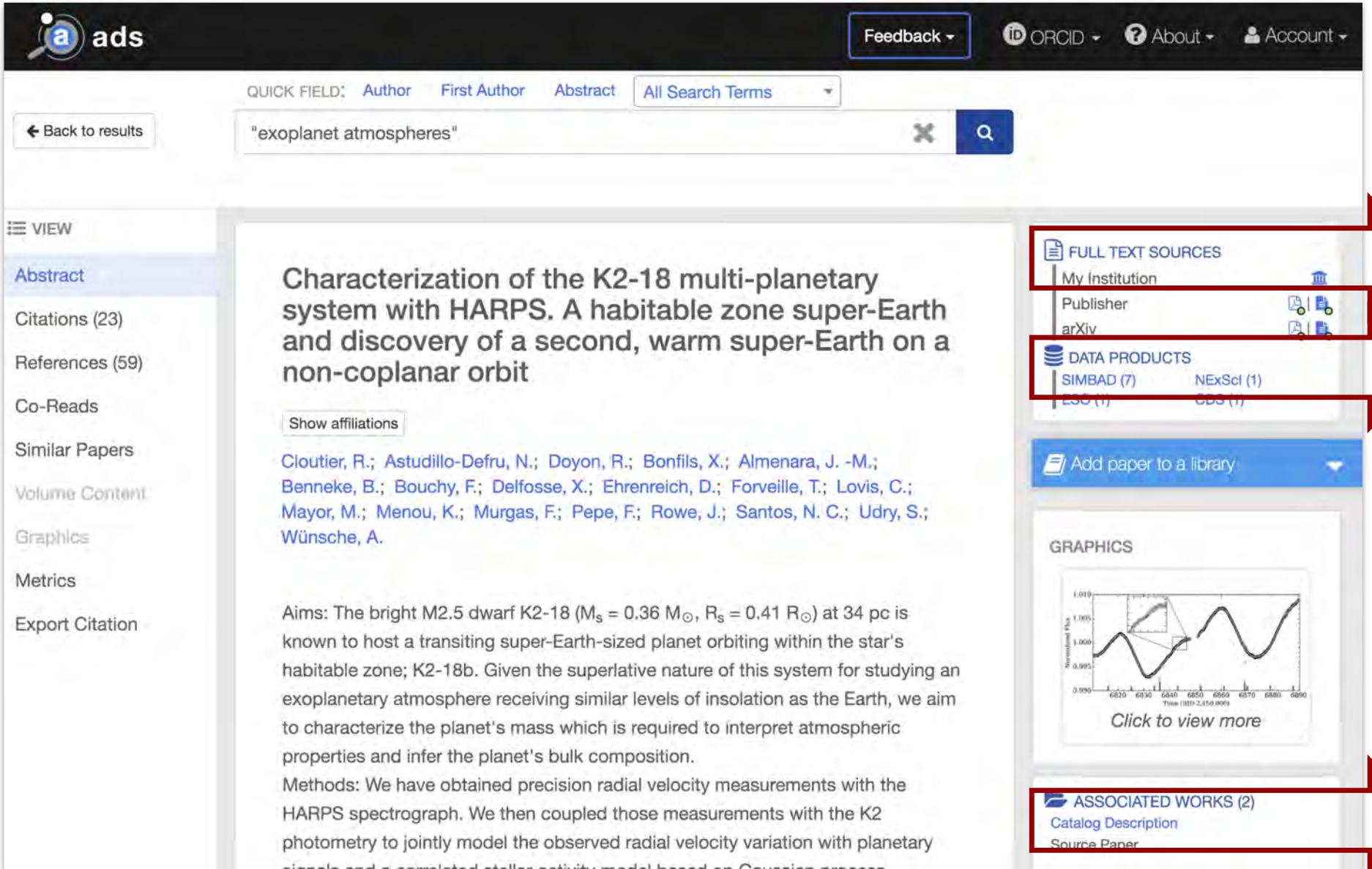
- Complete literature coverage: refereed journals, books, conferences, reports, PhD thesis, the so called “gray literature,” complete citation coverage
- High level data products: substantial effort collaborating with outside groups to link to measurements, index observing and funding proposals, software packages
- Data links: mine fulltext, collaborate with archives to link papers in our database to raw and reduced data behind them



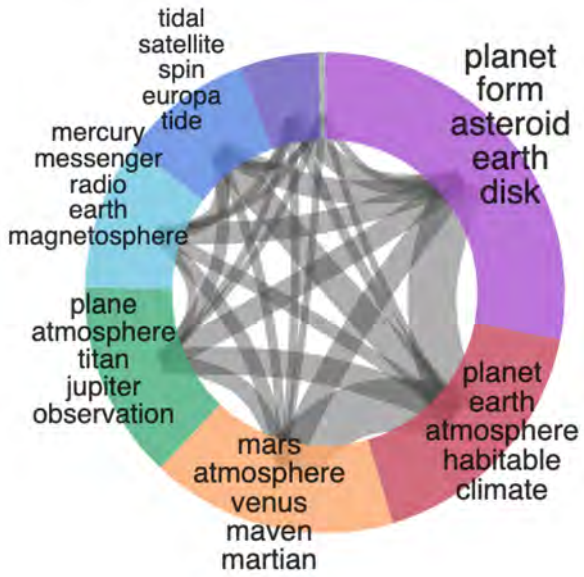
Making Data Discoverable...



… and Accessible



ADS now expanding its coverage to Planetary and Heliophysics literature & data

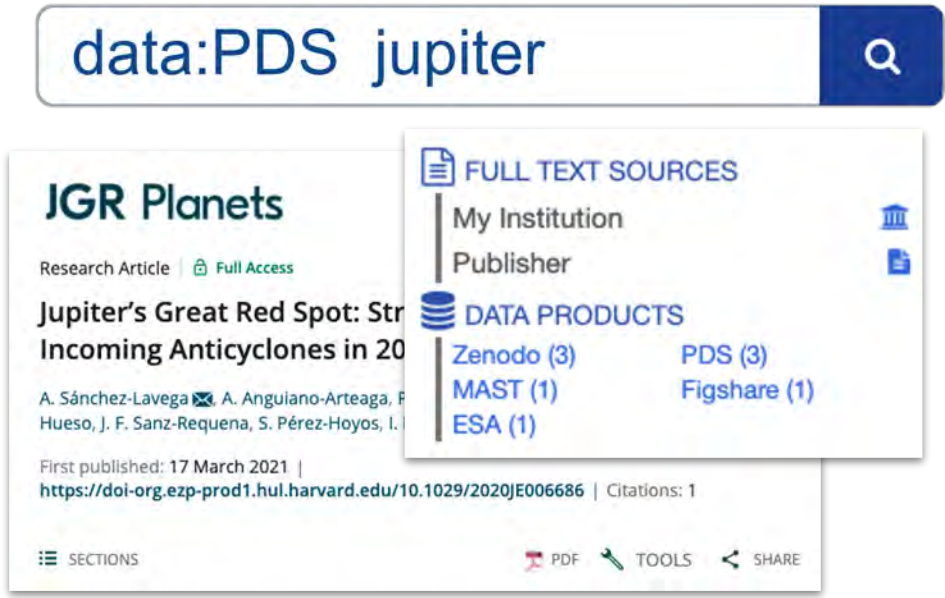


Literature

- Goal is for ADS to be as useful to PSD and HPD as it is to APD, providing current and accurate coverage of refereed and gray literature, preprints
- Effort has just begun, with main push over next two years

Software and data products

- Added 600 datasets from PDS SBN
- Added data links to 5.6K AGU journal articles
- Added links to 480 software packages cited in 513 AGU articles



More info: [DPS iPoster 507.07]

OpenScapes

Better Science for Future Us

We believe Open Science can accelerate data-driven solutions and increase diversity, equity, inclusion, and belonging.

We see many different pathways to Open Science, and have been intentional about building inclusion and sustainability into Openscapes as we support research teams and communities.

Erin Robinson & Julia Stewart Lowndes, Co-Directors

NASA Open-Source Science Workshop October 14, 2021

Artwork by Allison Horst

slides: https://nasa-openscapes.github.io/about

openscapes.org

@openscapes

NCEAS

METADATA GAME CHANGERS

Empower
research teams

Openscapes Champions

A mentorship program that helps researchers move from lonely science by exploring the open science landscape with their teams in a cohort of peers

Amplify leaders

"This isn't just about coding & GitHub, it's about changing the way we do science." - Dr. Malin Pinsky, Rutgers

"Openscapes has created a new way of thinking about merging empathy and science. That's an invaluable gift to me." - Dr. Halley Froehlich, UCSB

"Openscapes gave me a perspective of how all these open data science tools work together and can be used to bring natural resource conservation and ecology into the 21st century."
- Researcher, NOAA National Marine Fisheries Service

Biggest impact: research teams work more openly together

Reframe data-intensive research as collaborative effort not an individual burden.

students participate in research faster • grant money goes further • co-creating norms promoting diversity, equity & inclusion • new collaborations • leadership

Biggest lesson: power of research teams & data science to normalize open science

Engage Mentors

Develop Champions Mentor – professional development and leadership skills

Invest in Champions Program curriculum: NASA Earthdata cloud-specific materials

Attract research teams interested in better practices for data-intensive science

Empower research teams

Deliver Champions Program for a Cohort of research teams

Transform research teams workflows towards kinder, inclusive open

Inspire broader scientific communities through visible examples and leaders – Open science shift

Amplify leaders

Engage Mentors

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

- **People:** Building community, trust, normalizing talking about tutorials
- **Process:** Identifying common parts and shared needs for NASA Earth Science Cloud workflows
- **Technology:** shared skills & practices to co-create tutorials:
  - Carpentries Instructor Training
  - GitHub
  - 2i2c
  - Quarto

https://nasa-openscapes.github.io

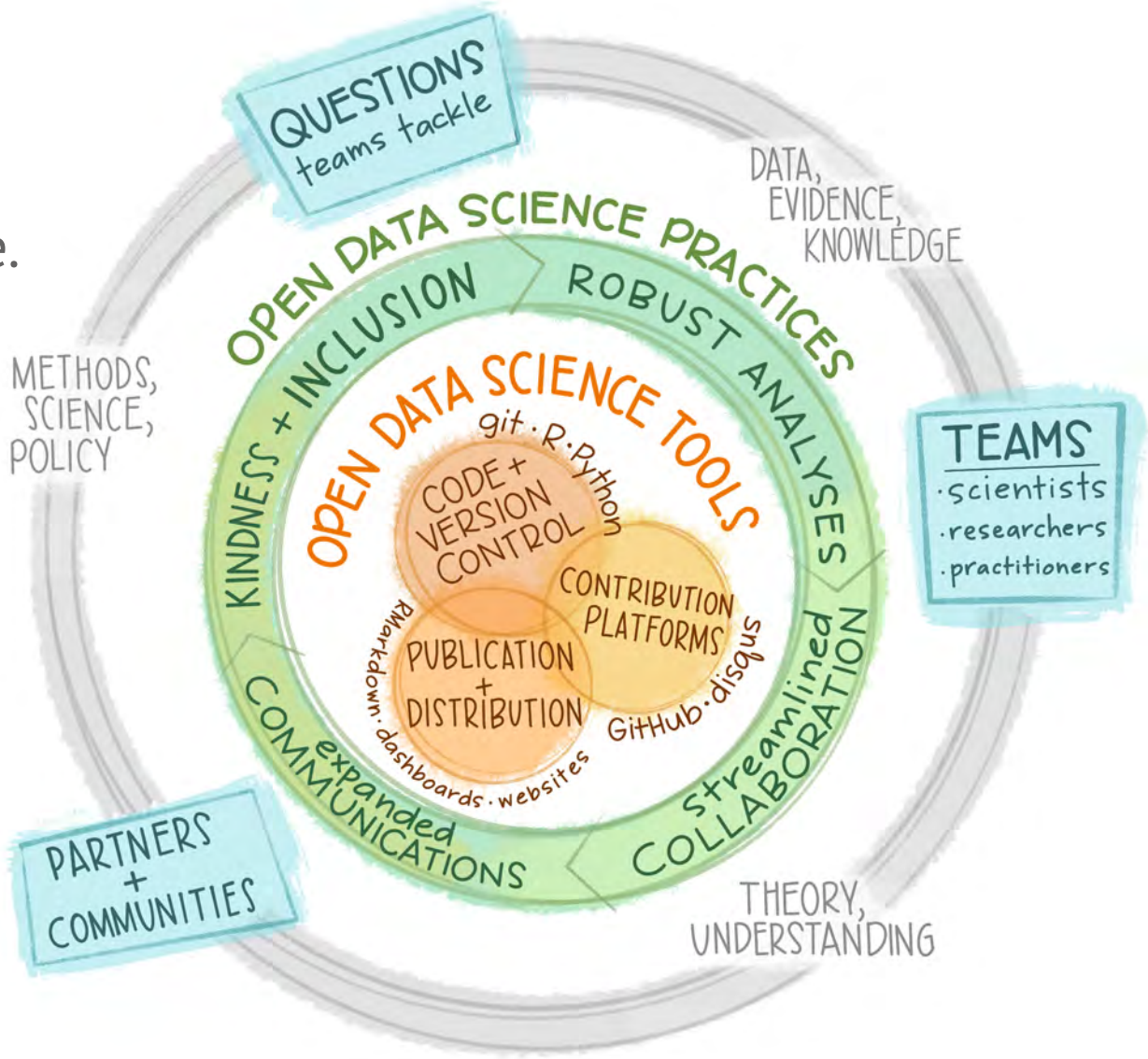
Openscapes approach

- **Researcher-centered, focused on teams**. Practice and feel safe working openly with yourself and your team; then ease into more.

- **Create space & place to explore & learn.** Cohort Calls, Seaside Chat, Co-Working; GitHub, R, Python, Google Drive, Slack; Efficiency Tips & Inclusion Tips.

- **Cultivate relationships & real connections.** Welcoming folks with diverse backgrounds; meeting where they are; skills to empower immediate work; kinder science.

- **Learning & iterating, openly.** Not a checklist but a continual practice.



https://openscapes.org/approach

Thank you

We're looking forward to working together!

Erin Robinson, MSc
Co-Director, Openscapes
Metadata Game Changers
erin@metadatagamechangers.com
@connector_erin

Julia Stewart Lowndes, PhD
Co-Director, Openscapes
National Center for Ecological Analysis and Synthesis (NCEAS), UC Santa Barbara (UCSB)
lowndes@nceas.ucsb.edu; @juliesquid

Join us:

Twitter: @openscapes

Web: openscapes.org

December: Community Call

Early 2022: NASA Openscapes Champions
nasa-openscapes.github.io

Further background:

Entryways to open data science and the power of welcome
Lowndes 2020, plenary at Earth Science Information Partners (ESIP) Meeting

Putting Data to Work
Robinson 2020, Leptoukh Lecture at the American Geophysical Union (AGU) Conference

Openscapes: Better Science for Future Us
Lowndes, Robinson 2021, plenary at SORTEE Meeting

Better Science for Future Us: Supporting NASA Earth science research teams' migration to the cloud with Openscapes
Robinson, Lowndes 2021, Pangeo Showcase, June 2021

slides: https://nasa-openscapes.github.io

Openscapes artwork by Allison Horst

Break Time
See you in 10 minutes!

Transforming to OPen Science

to change everything, we need everyone

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

> WE HAVE LOST THE LUXURY OF TIME
> COVID-19, CLIMATE CHANGE, ...
> THE WORLD IS CHANGING RAPIDLY
> SCIENTISTS MUST ADAPT

"We need every solution and every solver. As the saying goes, to change everything, we need everyone. What this moment calls for is a mosaic of voices-- the full spectrum of ideas and insights for how we can turn things around."

Ayana Elizabeth Johnson and Katharine Wilkinson (Eds.). All We Can Save: Truth, Courage, and Solutions for the Climate Crisis. 2021.

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

Why are we here?

We have to tackle a really hard problem: ***changing the cultural norms that are preventing us from embracing new ideas, truly working together and moving forward.***

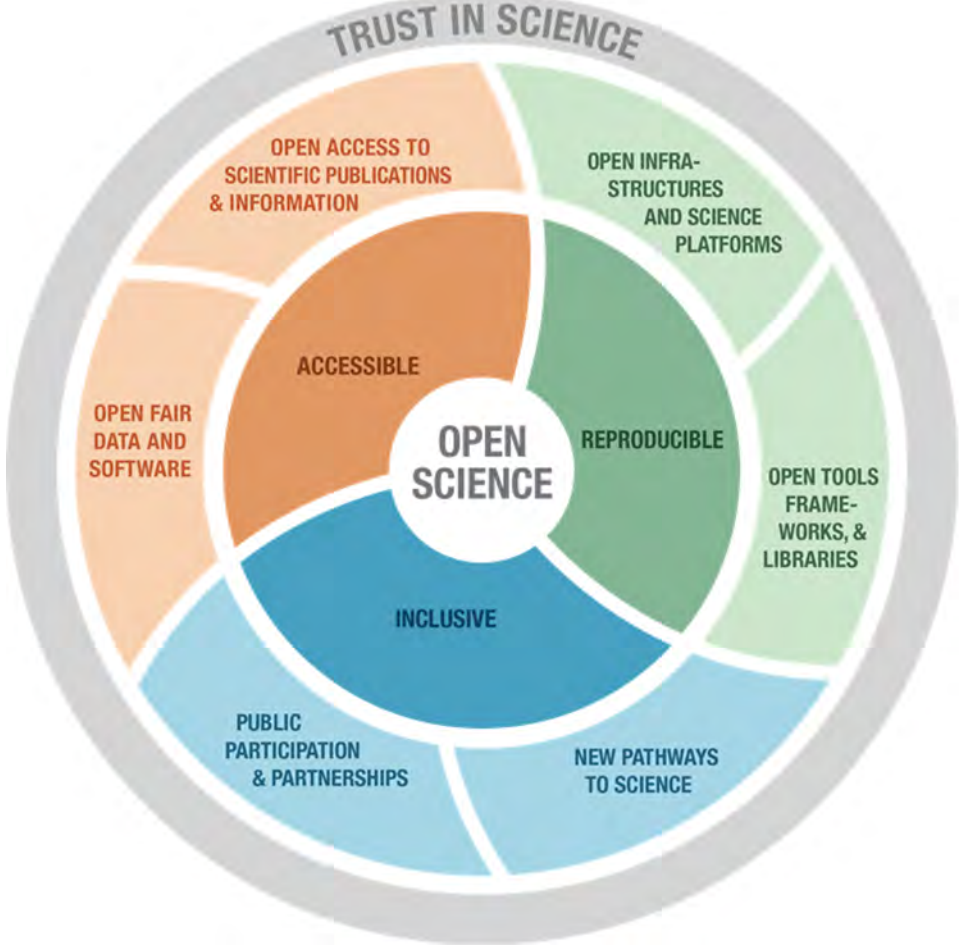
It isn't enough to talk about diversity and strengthen policies that foster inclusivity - we must change the power dynamics that disempowered and excluded people

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

What is open science?

Open Science Establishes Continued Trust in Science :

- Accessible: open data, open software, open information
- Reproducible: Make sharing and collaborating more efficient by supporting open software tools, frameworks, libraries, and open infrastructures
- Inclusive: innovative pathways to participation and expand public/private partnerships



Open knowledge ➡ Better data ➡ Better science ➡ Bigger impacts

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

Why do open science?

How:

- Open, transparent, collaborative, and inclusive scientific practices
- More accessible & verifiable scientific knowledge subject to scrutiny and critique

Results:

- Increases trust in science
- More efficient enterprise
- Improves quality
- Improves reproducibility
- Expands the impact of science
- Provides robust evidence for decision-making and policy
- Creates new pathways for participation
- More equitable

Sharing hidden knowledge

Image credit: Twentieth Century Fox

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

Why transform now?

Current challenges:

- Climate change
- Protecting our interconnected world from extreme space weather events
- Identifying threats from interplanetary space
- Searching for life beyond Earth
- Unlocking the secrets of the Universe

What are we going to do about it?

- Recognize the transformative potential of open science to reduce inequalities AND advance science

Image credit: NOAA

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

How is **NASA's Science Mission Directorate** going to respond?

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

- PROTECTING & IMPROVING LIFE ON EARTH
- LIFE ON OTHER PLANETS
- MYSTERIES OF THE UNIVERSE

Accelerating Scientific Discovery

These activities are designed to ***support and strengthen*** other NASA SMD initiatives on Inclusion, Diversity, Equity, and Accessibility (IDEA) and work for environmental justice.

Overview

- TOPS 5-year initiative will act as a catalyst to ***jump-start*** a suite of coordinated activities designed to rapidly transform science
- Designate **2023 as the Year of Open Science (YOOS)**

Goals

- Promoting a **common understanding** of open science, associated benefits and challenges, as well as diverse paths to open science.
- Investing in human resources, education, digital literacy and **capacity sharing** for open science.
- Fostering a **culture** of open science and aligning incentives for open science.
- Promoting innovative **approaches** for open science at different stages of the scientific process.
- Promoting international and multi stakeholder **cooperation** in the context of open science and in view of reducing digital and knowledge gaps.

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

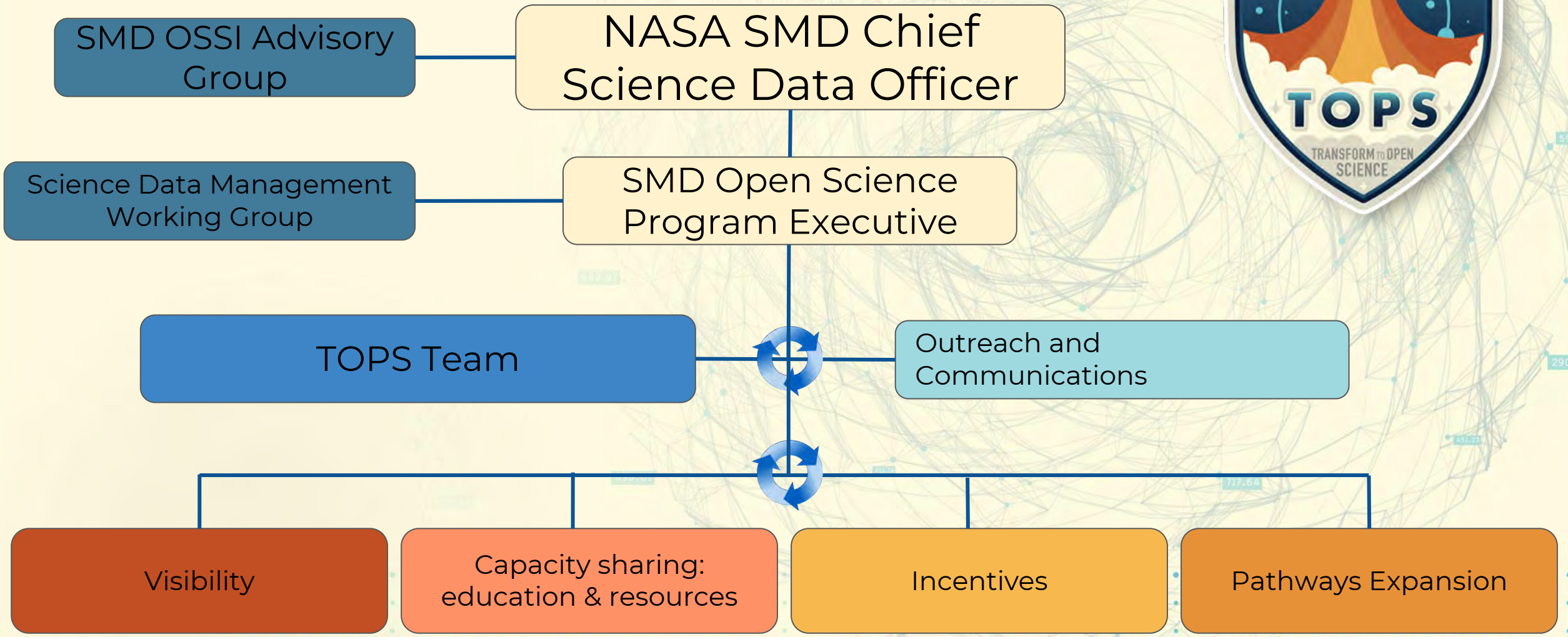
Key Performance Indicators for TOPS

In 5 years, TOPS will:

1. Increase understanding and adoption of open science principles and techniques in our Mission and Research Communities
   a. 75% of mission and research principal investigators certified in open science principles
   b. 20K scientists achieve open science certification
2. Accelerate major scientific discoveries through supporting the adoption of open science
   a. One major scientific discovery using open science methods supported in each division (5 community moon shots) within 5 years
3. Broaden participation by historically excluded communities
   a. Double participation by historically excluded communities in submitted proposals, applications from students, and participation in mission teams.

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

TOPS organizational structure



SMD = Science Mission Directorate

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

Areas of Action

Visibility

A lack of high-level support has left many scientists unsure about whether they are even 'allowed' to share knowledge and how moving to open science may impact their careers and funding.

- Promote 2023 - Year of Open Science (YOOS)
- Publish articles about TOPS and open science
- Build partnerships with scientific organizations to hold open science learning events at their annual meetings
- Provide visibility to solutions that advance adoption of open science
    - Eg. Software Release Agreements / tenure evaluations / award criteria
- Highlight open science success stories

Areas of Action

Capacity sharing: Learning resources and activities

Goal: 20K scientists & 75% PIs move to open science practises.

- Interactive open science platform populated by curated content that can be taught in-person or remote.
  - Build on existing resources to advance literacy in open source science methods, data science, tools and practices
    - Open data science events at annual meetings
    - Dedicated open data science summer schools
    - Open science cohorts
    - Open science events throughout year
    - Massive open online courses (MOOCs)

Areas of Action

Incentives

Reward and recognize 'the work'

- NASA Open Source Science Awards program
- Open science certifications / badges
- Prizes and challenges and cross-division science use cases, (eg. SpaceApps type events)
- Open science activities recognized in NASA reviews
- Support opening domain-specific course materials
- Support open learning resources
- Support to attend TOPS activities and events

Areas of Action

Pathways expansion

Double participation by historically excluded groups. Prioritize true change, collaborate with excluded communities to co-develop opportunities.

- New resources and opportunities for learning and participating in open science
  - Expand accessibility to free and open science research infrastructures
  - Leverage Public Participation and Partnerships
  - Host environmental justice targeted data science events
  - Invest in summer schools
- Engagement with historically excluded communities
  - Building partnerships with HBCU/MSI/HSI/TCU
    - Opening up hidden knowledge
    - Collaborative resource development
  - Funding (with mentorship) to attend TOPS activities and science meetings
  - Strengthen support for English as Second Language learners

to change everything, we need everyone

A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

Next Steps

> Prepare for 2023 YOOS
> Visibility
> Co-develop learning resources
> Incentives program
> Expand pathways & build partnerships
> Support open science
> Plan open science events

https://github.com/nasa/Transform-to-Open-Science

Breakout Rooms

Breakout Rooms
Session 1

During the breakout sessions, attendees may select their room of choice for each session. Sessions will be 20 minutes. The moderator will give a 5 minute and then 1 minute warning to summon the group back to the main room. Each breakout will have their own Jamboard exercise. Topic Areas include:

- **Breakout Room #1: Open Source Science Initiative: General**
- **Breakout Room #2: Transform to Open Science: Training**
- **Breakout Room #3: Policy**
- **Breakout Room #4: AI/ML**
- **Breakout Room #5: SMD Data Catalog**

In breakout rooms, attendees are encouraged to spark open discussion involving the room's selected topic. Moderators will be present to acknowledge any questions, or requests involved in the open conversation. This is YOUR chance to make heard YOUR questions, comments, or concerns amongst your peers.

Breakout Rooms
Session 2

During the breakout sessions, attendees may select their room of choice for each session. Sessions will be 20 minutes. The moderator will give a 5 minute and then 1 minute warning to summon the group back to the main room. Each breakout will have their own Jamboard exercise. Topic Areas include:

- **Breakout Room #1:Open Source Science Initiative: General**
- **Breakout Room #2: Inclusion and Open Source Science Initiative**
- **Breakout Room #3:Transform to Open Science: Analysis ready data for the cloud**
- **Breakout Room #4: Transform to Open Science: Platforms**

In breakout rooms, attendees are encouraged to spark open discussion involving the room's selected topic. Moderators will be present to acknowledge any questions, or requests involved in the open conversation. This is YOUR chance to make heard YOUR questions, comments, or concerns amongst your peers.

Wrap Up

Next Steps

- Join the TOPS Mailing list
  - Let us know what you are doing!
  - Let us know where you are interested in joining!
- Reply to SPD-41 Request for Information once it is released
  - How will the proposed changes to the existing information policy impact the research activities of your communities?
  - What support, services, training, funding, or further guidance is needed to support the successful implementation of the existing or proposed information policy?
- Reach out to have us come speak about OSSI or TOPS to your organization or group.

Next Steps

Add your thoughts to the Jamboard or in the chat for what's next!

- What should be the next meeting or meetings be like and when?
- What would be useful besides meetings?

Wrap Up

Thank you all for joining!

Recordings and relevant links to the workshop will be provided on the workshop website following the meeting.

Pertinent workshop questions regarding the presentation can still be submitted to the IO Tool and assessed at this link: https://nasa.cnf.io/sessions/w411/#!/dashboard

Closing Remarks